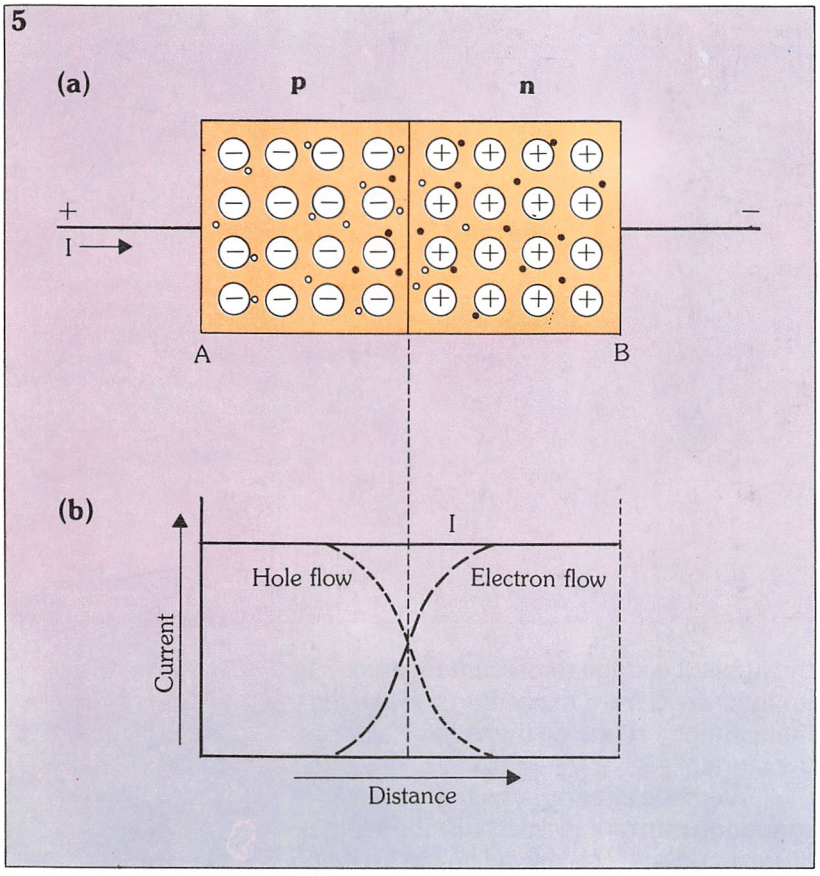
(continued from part 5)

We know that a voltage applied to a reverse biased p-n junction brings about an increase in the potential barrier. This stops all flow of majority carriers and as the minority carrier current is unchanged, the total current is negligible. Let's see what happens when a forward bias is applied.

Forward biased p-n junctions

Figure 5 shows a forward biased p-n junction. Here the positive terminal of the battery is connected to the p-type section and the negative terminal to the n-type.

When a voltage is applied electrons from the negative terminal are pumped into the n-zone: these repel the free electrons in this area, forcing them towards the p-n junction. At the same time bound electrons are being withdrawn from the p-zone, creating new holes. The new holes repel the old holes, moving the holes towards the p-n junction. So the holes in the p-type silicon and the free electrons in the n-type are moving towards each other.



5. The p-type semiconductor is connected to the positive terminal of a supply and the n-type to the negative. The relationship between the flow of electrons and holes at the junction is shown.

When the holes and free electrons meet at the junction the free electrons fall into the holes, thereby neutralizing some donor and acceptor ions and reducing the potential barrier. So far the current has only increased slightly; any additional increase in the current will reduce the potential barrier still further, so that the majority carriers can move freely across the junction.

As they cross over, electrons combine with holes in the p-zone and holes combine with electrons in the n-zone. In this way, a current flows across the junction which will increase rapidly with even a small voltage increase.

The current which flows when a forward bias is applied is called a **forward current.** It is, in essence, a flow of majority carriers across the junction. But it's important to remember that once these have crossed the junction they become minority carriers on the other side.

Now let's look at the nature of the current flowing from electrode A to electrode B through the p-n junction. In the p-zone, near electrode A, it consists of a flow of holes. In the n-zone, around electrode B, it consists of a flow of electrons. In the vicinity of the junction the current consists of a combination of both electrons and holes. The proportions change gradually from all holes in the p-zone near electrode A, to all electrons in the n-zone near B. This is illustrated in *figure 5b.*

As soon as the holes in the p-zone approach the junction some will meet electrons which have crossed over from the n-zone and recombination will begin to take place. As the holes progress into the n-zone this combination will become faster and faster until all the holes recombine, disappearing completely within a certain distance.

Likewise the flow of electrons in the n-zone will begin to diminish due to recombination in the vicinity of the junction. They will disappear at a certain distance in the p-zone. Naturally the total current must be equal to the external current I at all points within the structure.

It can be seen that the current for a certain distance on both sides of the junction is partly composed of minority

carriers which have crossed the junction. The phenomenon which causes majority carriers to cross the p-n junction and flow in the other zone as minority carriers is called **injection.** A stream of holes has been *injected* into an n-type region where the current is composed of electrons, and vice versa.

The importance of injection will become clear when we deal with more complicated semiconductor structures. Up to now we have been looking at a symmetrical structure using n and p-types with the same concentration of electrons and holes. In practice a semiconductor is often constructed so that the concentration is greater on one side than the other. This will be covered in *Solid State Electronics – 9.*

Electrical characteristics of a p-n junction

Now that we have looked at the different characteristics of the forward and reverse bias of a p-n junction, let's consider its overall electrical characteristics. A typical voltage/current curve of a p-n junction is shown in *figure 6.*

When the forward voltage is increased from zero only a little direct current will flow. But when the voltage is only a little lower than the potential barrier the current increases rapidly for each small voltage increase; the forward resistance is therefore low.

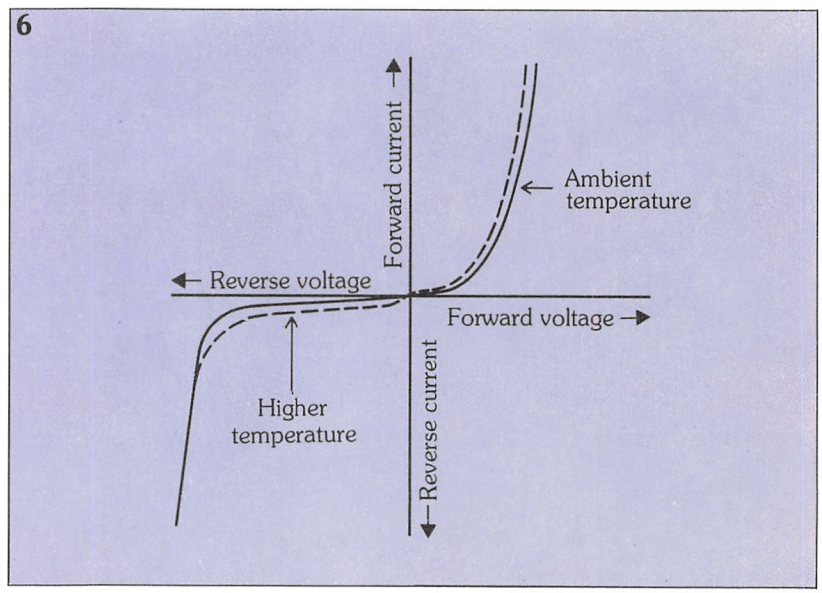
In the opposite direction, the reverse current remains constant at a low level of reverse saturation current, even if the voltage increases. This reverse saturation current is very low – about 100 nA for a typical transistor – so in fact the reverse resistance is extremely high. This characteristic enables a p-n junction to rectify alternating current into direct current.

However, there is a limit to the reverse voltage which the material can tolerate. Should the reverse voltage reach a strength known as **reverse voltage breakdown,** the reverse current then increases a great deal for each small voltage increase. As we know, the reverse voltage forces the minority carriers, which constitute the reverse saturation current, to cross the junction at a speed which depends on the strength of the voltage and the mobility of the carriers.

We also know that the number of thermally generated carriers is limited. But if the voltage applied is great enough the carriers will move so fast that they will collide with the crystal atoms and break the covalent bonds. We now have additional charge carriers which can in turn pick up enough speed to break the bonds with other atoms. And so the process escalates. As a result, the current increases considerably with each small voltage increase. This is called **avalanche breakdown,** for obvious reasons.

When reverse breakdown takes place in a p-n junction the reverse current can become strong enough to damage the device. But the problem can be overcome by using a suitable resistance in series with the applied voltage to prevent the maximum current from exceeding the danger limit, thereby reducing the voltage across the diode.

6. The voltage/current curve for a typical p-n junction at ambient temperature is shown by the solid line. The dotted line indicates the change at a higher temperature.



We have already seen that reverse saturation currrent increases as the temperature increases. The dotted line in *figure 6* illustrates this effect. You will see that the forward current also increases slightly with the temperature, due to an increase in the concentration of carriers.

How p-n junctions are used

The p-n junction forms the basis of most semiconductor devices. A single p-n junction makes a diode, as we know. Transistors and thyristors, among others, use

p-n-p, n-p-n and p-n-p-n combinations. We shall cover these devices in greater detail in later chapters. The actions of the junctions can be varied by changing the levels of doping and even the dopants used. However, in all cases the action of a p-n junction is basically the same as described here.

Glossary

acceptor atom	atoms with three valency electrons, used to dope pure semiconductor material to give p-type semiconductor
avalanche breakdown	the breakdown which occurs in a reverse biased p-n junction when some of the carriers gain enough energy to liberate new electron/hole pairs by impact ionization, which in turn can generate further pairs
depletion layer	this is the space charge region in a semiconductor that has a net charge due to insufficient mobile charge carriers. A depletion layer is inevitably formed at a p-n junction in the absence of an applied field
diffusion current	flow of electrons and holes across a p-n junction which is not dependent on an applied voltage
donor atom	donor atoms with five valency electrons, used to dope pure semiconductor material to give n-type semiconductor
drift current	flow of charge carriers across a p-n junction due to an applied voltage
forward bias	electrical conditions at a p-n junction when the anode is at a more positive voltage than the cathode
negative ion	an atom, or molecule, that has a negative electric charge. Negative ions (anions) contain more electrons than are necessary for electrical neutrality
p-n junction	the dividing line in a semiconductor between a p-region and an n-region
p-n junction diode	semiconductor device which permits electrons to flow through it in one direction only
positive ion	an atom, or molecule, that has a positive electric charge. Positive ions (cations) contain fewer electrons than are necessary for electrical neutrality
potential barrier	opposition to the movement of majority carriers which develops automatically in the space charge area
reverse bias	electrical condition at a p-n junction when the cathode is at a more positive voltage than the anode when reverse voltage is applied
reverse saturation current	reverse current of minority carriers which diffuse across the depletion layer under a reverse voltage
space charge region	area of the p-n junction which carries a charge. See depletion layer

Diodes: how they work

Diodes and current flow

It's important at this stage to clarify the difference between conventional current flow and electron flow. You will have noticed that in diagrams the conventional current is shown to flow in the opposite direction to the electron flow.

Figure 1 shows the symbols for a diode and an n-p-n transistor. The diode allows the electrons to flow in the opposite direction to the direction of the solid arrow. Likewise, the transistor allows the electrons to flow from emitter to collector when electrons are withdrawn from the base, but the current arrow points the other way.

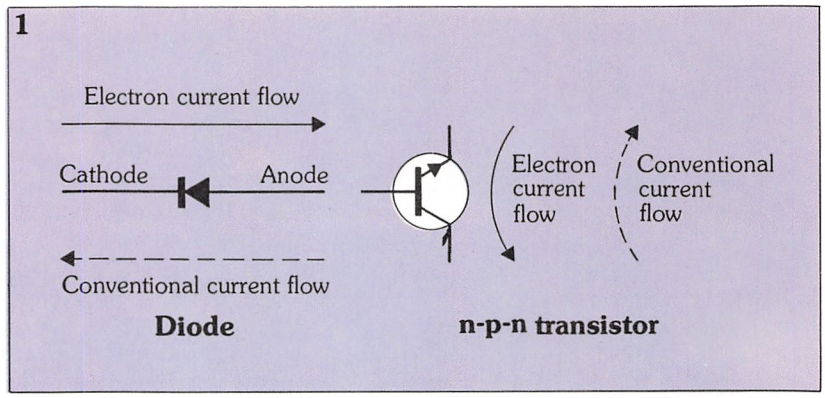
In both symbols the solid arrow indicates the direction of the conventional current flow.

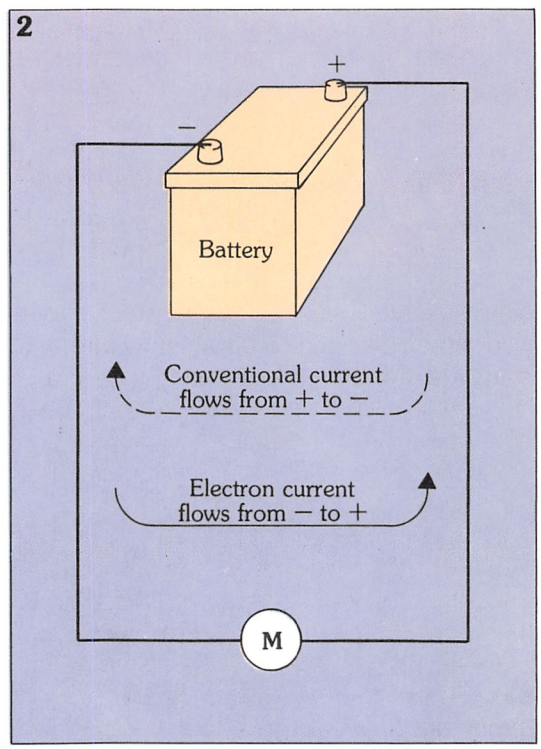
What is conventional current?

Benjamin Franklin (1706-90) devised the concept of 'conventional current' flow, basing his judgement on the results of his experiments with static electricity. In Franklin's time very little was known about the phenomenon of electricity and he described it as an 'invisible fluid' which all substances contained. In statically charging and discharging bodies, he reasoned that this fluid was transferred from one to the other.

He termed an excess of fluid a positive charge. His theory said that the fluid tended to flow from a positively charged substance to a negatively charged one. But he didn't know which bodies possessed the positive or negative charges, and so he guessed and categorised materials accordingly.

Franklin's explanation became accepted as convention in all areas of electrical theory. *Figure 3* shows how, according to Franklin, a conductor can possess any voltage, negative (fluid deficiency) or positive (excess of fluid).



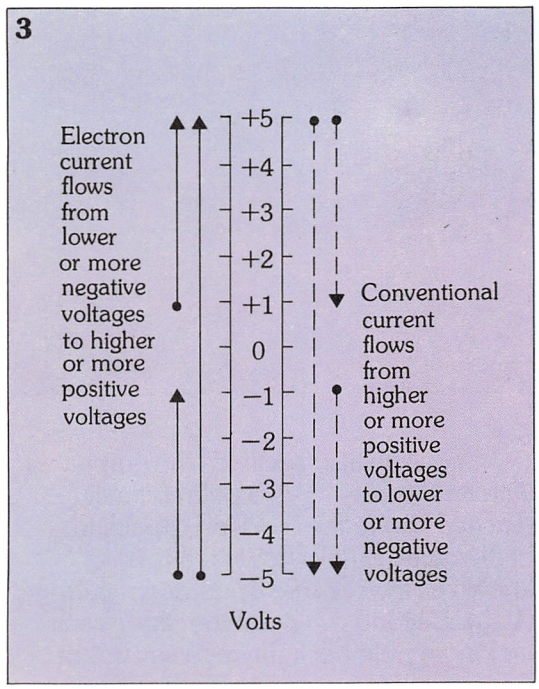


1. Conventional symbols for a diode and an n-p-n transistor.

2. Direction of conventional and electron current flow.

A century later, scientists discovered that the 'electric fluid' flowed from the negative to positively charged bodies, and later defined electricity as a flow of electrons. So, in fact, electrons flow in the opposite direction to conventional current.

To make this discovery fit in with Franklin's theory (which by now had worldwide acceptance) electrons had to be

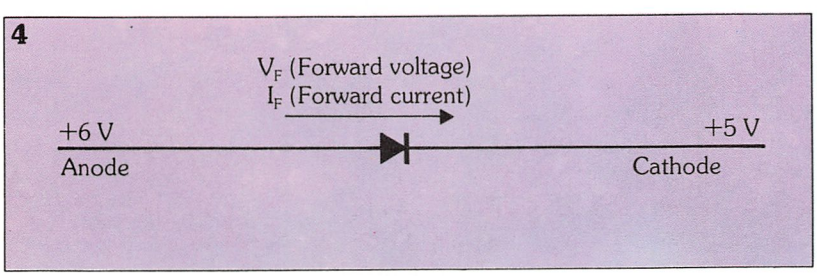


3. Electron current flow is in the opposite direction to Franklin's conventional current.

defined as negative, as the direction in which they flow is the reverse of conventional current.

Remember, conventional current is in effect the flow of an imaginary fluid, but to redefine Franklin's theory would have been expensive and confusing, as it had been accepted and used for over a hundred years. Conventional current flow however is still a valid concept as long as you remember that electrons flow in the opposite direction. Electrical circuits and machinery are designed with conventional current flow in mind. But *electron* flow does become important when looking at the intricate workings of semiconductors and other such devices.

In this text conventional current terms are always used, except when the internal functioning of components is being explained. Then, electron current terminology is used, as it provides the clearest explanations of the workings of semiconductors.



4. A diode will allow current to pass in the forward direction, but blocks it in the reverse direction.

The behaviour of diodes

So far we know that a diode lets current pass in the forward direction and blocks current in the reverse direction. This can be seen in *figure 4*. (Note that the direction used here is that of conventional current.)

The **behaviour** of a diode is the relationship between the voltage (V) and the current (I). The **forward voltage (V_F)** is the amount by which the anode voltage is greater than the cathode voltage. (Remember that this is positive voltage.) In our example the forward voltage is 1 V since the anode voltage (+6 V) exceeds the cathode voltage (+5 V) by 1 V. The **forward current (I_F)** is simply the amount of current at a given forward voltage.

Since no diode is perfect, current is not always completely blocked in the reverse direction. When a **reverse voltage (V_R)** is applied, a small amount of **reverse current (I_R)** flows. The reverse voltage is the amount by which the cathode voltage exceeds the anode voltage. Reverse current is the amount of current at a given reverse voltage.

How diode behaviour is illustrated graphically

The curve in *figure 5* is the key to understanding diode specifications; nearly all the important diode characteristics are represented by this typical behaviour curve. Once you can read a diode graph, you can understand its particular specifications at a glance.

The top section of the vertical co-ordinate represents the forward current (I_F); the lower section represents the reverse current (I_R). The horizontal coordinate shows the forward voltage (V_F) on the right and the reverse voltage (V_R) on the left.

Notice that the relationship of the scale changes considerably between the forward voltage and the reverse voltage: 1 V of V_F on this graph is equal in distance to −50 V of V_R. This is because the critical voltages in the forward conducting mode are far lower than those in reverse.

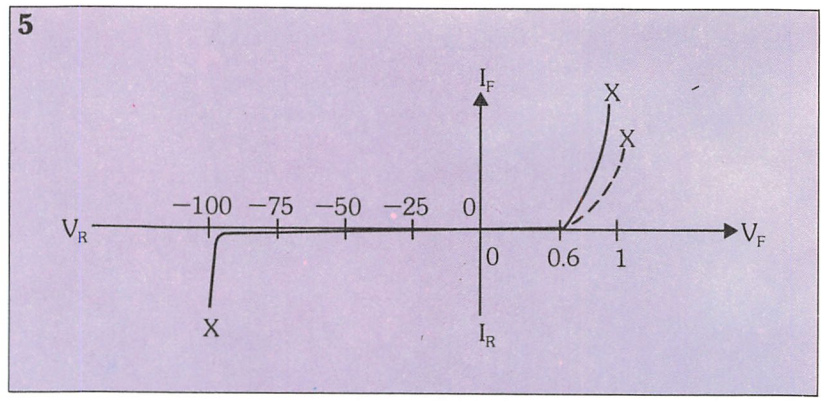
Let's see what the graph tells us about a diode when a forward voltage is applied. The axes divide the graph into four quadrants. The top right quadrant shows the behaviour of the diode under

forward voltage conditions. Looking at the solid line, we see that when the voltage is zero the forward current is also zero. As the forward voltage increases, the current increases – gradually at first and then more rapidly. The curve makes a rather sharp bend or **knee** when the voltage reaches a certain strength (0.6 V in silicon diodes). The knee indicates that the current begins to increase dramatically at this voltage, and the curve swings upwards. This point can be considered as the **threshold voltage** where the diode really begins to turn on (in germanium diodes the threshold voltage is about 0.3 V).

From the knee on, a slight increase in voltage results in a much greater current increase. The upward curve ends abruptly at a limit where the diode burns up because the heat generated is more than the device can dissipate. Remember that by Ohm's law, power dissipation (watts) is equal to current (amps) multiplied by the voltage (volts). So as the current and voltage increase, more and more heat is generated until eventually the diode burns up (point X).

Every diode has a forward conduction curve which is similar to the one illustrated in *figure 5*. It is the slight differences in the curves and the burnout points which distinguish different types. Another diode might have a behaviour pattern like the one represented by the broken line curve. We know from the curve that this diode conducts less current for a given voltage increase and that it can handle less power than the first one because the burnout point and hence the power rating is lower.

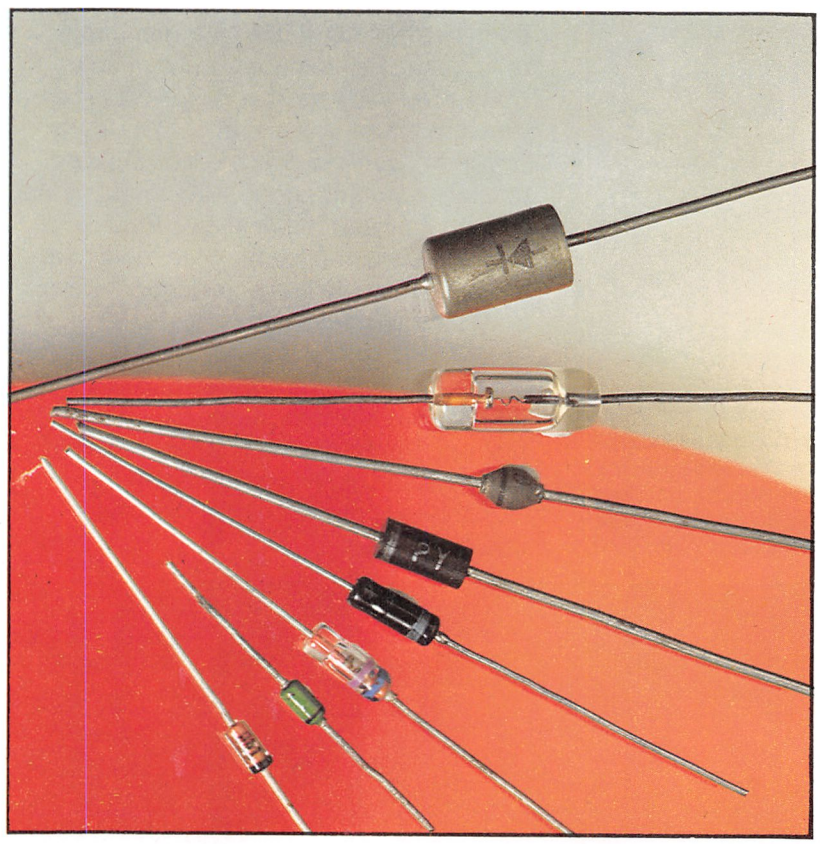
Now let's look at the bottom left quadrant to see how the diode behaves under reverse voltage conditions. With the increase in reverse voltage, the increase in reverse current is as shown. You can see that up to a certain point only a very small current 'leaks through' as the reverse voltage increases. So the diode is quite effective as a one-way valve – although not perfect. In this typical example, increasing the forward voltage from 0.6 V to just less than 1 V produces a very large increase in current. But under a reverse voltage condition, any voltage from 0 V to −75 V produces a negligible amount of current.



5. A diode characteristic curve.

You can also see from the graph that with an increase in V_R a point is finally reached where the blocking capability of the diode breaks down – this is known as the **reverse breakdown voltage (V_{BR}).** Beyond this point the diode can no longer hold back the reverse current. The current increases dramatically and the diode soon burns up. This happens very quickly because at these relatively high voltages only a small amount of current is enough to generate a much higher power which can destroy the diode.

A selection of different types of diode.



The most important diode specifications

So how much does the graph tell us about the most important diode specifications and how much does it leave out?

There are five really important details we need to know:

I_F (forward current) – the amount of current the diode can handle without burning up; when combined with the forward voltage it gives us a measure of how much power the device can dissipate.

V_F (forward voltage) – the voltage level across the diode when forward biased i.e. conducting.

I_R (reverse current) – the amount of current which leaks through the diode at various reverse voltages.

V_{BR} (reverse breakdown voltage) – the reverse voltage beyond which the current begins to rise very rapidly.

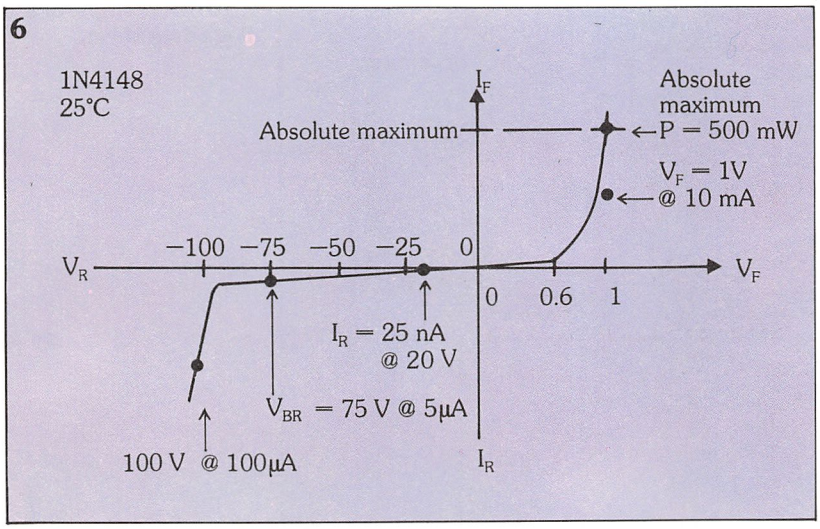
T_{RR} (reverse recovery time) – the time it takes a diode to recover from forward conduction before it begins to block reverse current. This time becomes important when alternating currents are considered. The higher the frequency of alternating current applied to the diode, the quicker it must respond to rectify it.

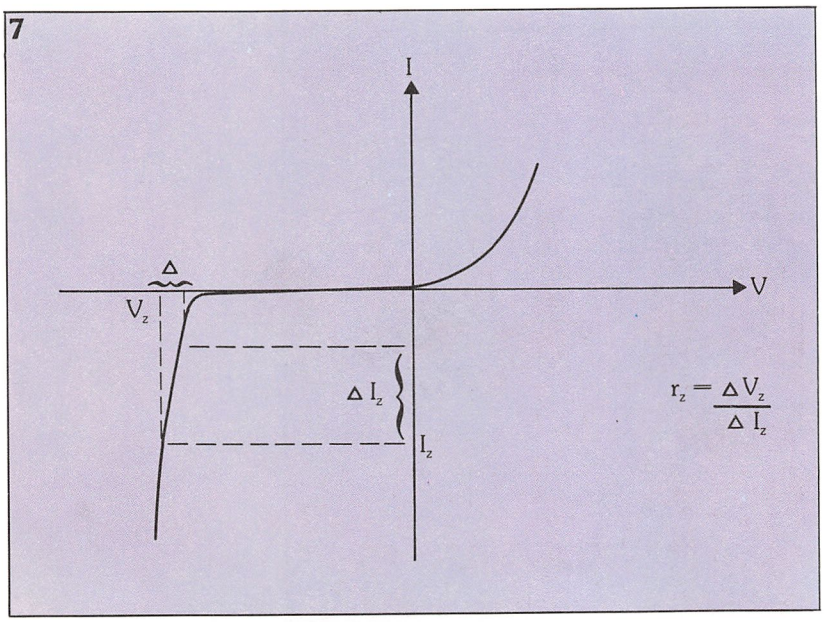
All of these specifications, with the exception of T_{RR}, can be read directly from an I-V curve – that is, a current-voltage curve of the kind we have just been looking at. *Figure 6* shows where each of these points is situated on a typical curve.

It is important to note that these results have been obtained by testing the device under fixed temperature conditions. This temperature is 25°C for the surrounding air. If the temperature were higher, 100°C say, there would be a much greater current for a given voltage drop because a higher temperature, as we know, causes an increase in both forward current and reverse leakage current. So, the same diode would have different curves at different temperatures.

6. A current-voltage curve from which various specifications can be read.

7. Voltage-current characteristics of a p-n junction. Note how the small increase in reverse voltage beyond the V_{BR} point causes a dramatic increase in current.





Reverse breakdown voltage

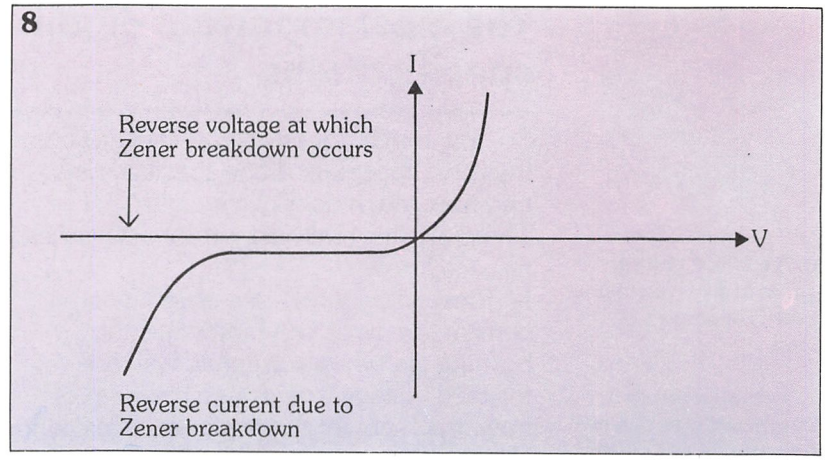
We have seen that the reverse breakdown voltage (V_{BR}) restricts the use of semiconductor diodes. *Figure 7* shows in detail how any small increase of reverse voltage beyond the V_{BR} point causes a dramatic increase in current.

Let's look again at the mechanism which causes the breakdown of a diode as the voltage increases. According to present theories, this phenomenon can have two different causes.

One of these is the already familiar avalanche breakdown, caused by minority carriers which are speeded up by the reverse voltage. When the voltage exceeds a certain level the minority carriers build up enough speed to collide with the covalent bonds and break them. This creates new hole/electron pairs which in turn collide with other bonds, and so the process continues. High reverse currents are created in this way and the diode is

threatened by total breakdown.

The second mechanism is the **Zener type breakdown.** The scientist Zener, noticed that leakage current remained very small for reverse voltages up to certain threshold values. When this value is exceeded there is a sudden rise in reverse current, known as the Zener current. This is due to the intense electric field at the junction breaking the covalent bonds in the semiconductor materials. This reverse current is not destructive so long as the power handling capacity of the diode is not exceeded (see *figure 8*).



8. Voltage-current characteristics of a Zener diode.

9, 10. An example of a diode data sheet.

UNITRODE†

TYPES 1N4148, 1N4149, 1N4446, 1N4447, 1N4448, 1N4449
PLANAR SILICON SWITCHING DIODES

BULLETIN NO. DL-S 667269, OCTOBER 1966 — TYPES 1N4148, 1N4149, 1N4446, 1N4447, 1N4448, 1N4449

- **Small-Size, Whiskerless, Double-Plug Construction**
- **Extremely Stable and Reliable High-Speed Diodes**

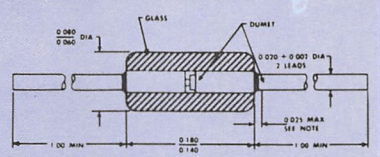
Electrical Equivalents

1N4148 • 1N914
1N4149 • 1N916
1N4446 • 1N914A
1N4447 • 1N916A
1N4448 • 1N914B
1N4449 • 1N916B

mechanical data

The glass-passivated silicon wafer is encased in a hermetically sealed glass package. High-temperature bond between wafer and leads insures integral positive contact under extreme environmental conditions.

*CASE OUTLINE



DIMENSIONS ARE IN INCHES
NOTE: WITHIN THIS ZONE DIAMETER OF EACH LEAD IS UNCONTROLLED

CATHODE END IS DENOTED BY COLOR BAND

***absolute maximum ratings at 25°C free-air temperature (unless otherwise noted)**

$V_{RM(wkg)}$	Working Peak Reverse Voltage	75 V
P	Continuous Power Dissipation at (or below) 25°C Free-Air Temperature (See Note 1)	500 mW
T_{stg}	Storage Temperature Range	−65°C to 200°C
T_L	Lead Temperature 1/16 Inch from Case for 10 Seconds	300°C

***electrical characteristics at 25°C free-air temperature (unless otherwise noted)**

PARAMETER		TEST CONDITIONS	1N4148		1N4149		1N4446		1N4447		1N4448		1N4449		UNIT
			MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	
$V_{(BR)}$	Reverse Breakdown Voltage	I_R = 5 μA	75		75		75		75		75		75		V
		I_R = 100 μA	100		100		100		100		100		100		V
I_R	Static Reverse Current	V_R = 20 V		25		25		25		25		25		25	nA
		V_R = 20 V, T_A = 100°C										3		3	μA
		V_R = 20 V, T_A = 150°C		50		50		50		50		50		50	μA
V_F	Static Forward Voltage	I_F = 5 mA									0.62	0.72	0.63	0.73	V
		I_F = 10 mA		1		1									V
		I_F = 20 mA						1		1					V
		I_F = 30 mA												1	V
		I_F = 100 mA										1			V
C_T	Total Capacitance	V_R = 0, f = 1 MHz		4		2		4		2		4		2	pF

NOTE 1: Derate linearly to 200°C at the rate of 2.85 mW/deg.
†Trademark of Texas Instruments
* Indicates JEDEC registered data

How to read a data sheet

We're now ready to look at a diode data sheet – an example is shown in *figures 9* and *10*. All diode manufacturers present their data in much the same way, so this one from Texas Instruments will serve as a general model.

Most of the data sheet is self-explanatory. The heading shows that this particular sheet covers an entire family of planar silicon switching diodes; this includes 1N4148 – which happens to be the diode whose I-V curve was illustrated in *figure 6* – plus five other very similar diodes.

Below the heading is a summary of the most important characteristics, followed by the mechanical data which consists of a description of the diode's structure and a technical drawing of the package.

Then follows a list of the absolute maximum ratings. These are conditions which cannot be exceeded without damaging the diodes and apply to all the diodes listed.

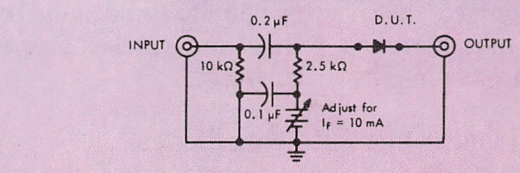
At the bottom of this page is a table of electrical characteristics which gives the different characteristics for each diode in a

TYPES 1N4148, 1N4149, 1N4446, 1N4447, 1N4448, 1N4449
PLANAR SILICON SWITCHING DIODES

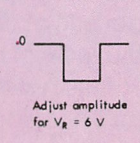
***switching characteristics at 25°C free-air temperature**

PARAMETER		TEST CONDITIONS	1N4148		1N4149		1N4446		1N4447		1N4448		1N4449		UNIT
			MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	MIN	MAX	
t_{rr}	Reverse Recovery Time	I_F = 10 mA, V_R = 6 V, i_{rr} = 1 mA, R_L = 100 Ω, See Figure 1		4		4		4		4		4		4	ns
$V_{FM(rec)}$	Forward Recovery Voltage	I_F = 50 mA, R_L = 50 Ω, See Figure 2										2.5		2.5	V

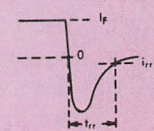
***PARAMETER MEASUREMENT INFORMATION**



TEST CIRCUIT



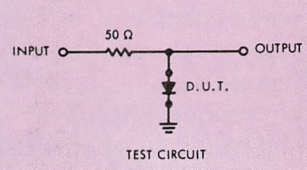
INPUT VOLTAGE WAVEFORM



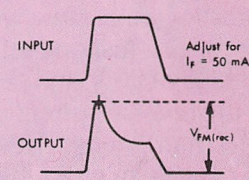
OUTPUT CURRENT WAVEFORM

FIGURE 1 — REVERSE RECOVERY TIME

NOTES: a. The input pulse is supplied by a generator with the following characteristics: Z_{out} = 50 Ω, $t_r \leq$ 0.5 ns, t_p = 100 ns.
b. The output waveform is monitored on an oscilloscope with the following characteristics: $t_r \leq$ 0.6 ns, Z_{in} = 50 Ω.



TEST CIRCUIT



VOLTAGE WAVEFORMS

FIGURE 2 — FORWARD RECOVERY VOLTAGE

NOTES: c. The input pulse is supplied by a generator with the following characteristics: Z_{out} = 50 Ω, $t_r \leq$ 30 ns, t_p = 100 ns, PRR = 5 to 100 kHz.
d. The output waveform is monitored on an oscilloscope with the following characteristics: $t_r \leq$ 15 ns, $R_{in} \geq$ 1 MΩ, $C_{in} \leq$ 5 pF.
*Indicates JEDEC registered data

TEXAS INSTRUMENTS
INCORPORATED
SEMICONDUCTOR-COMPONENTS DIVISION
POST OFFICE BOX 5012 • DALLAS, TEXAS 75222

PRINTED IN U.S.A.
TI cannot assume any responsibility for any circuits shown or represent that they are free from patent infringement.
TEXAS INSTRUMENTS RESERVES THE RIGHT TO MAKE CHANGES AT ANY TIME IN ORDER TO IMPROVE DESIGN AND TO SUPPLY THE BEST PRODUCT POSSIBLE.

separate column.

The second page of the data sheet (*figure 10*) has a table of switching characteristics. This gives more details about the performance of each type, operating within the absolute maximum ratings.

The rest of the page is devoted to parameter measurement information which provides a complete description of exactly how the tests were carried out on each diode. Armed with this information the purchaser is in a position to verify the specifications for himself.

How the data sheet relates to the I-V curve

We have seen how the most important diode characteristics, with the exception of reverse recovery time, were represented on the I-V curve. We'll now compare the graph in *figure 6* with the data sheet to demonstrate where all the important specifications can be found on the sheet.

P (power dissipation) can be found in the absolute maximum ratings table. It is 500 mW (milliwatts). Sometimes this rating is specified in terms of the absolute maximum forward current.

$\mathbf{V_F}$ (forward voltage) is found in the table of electrical characteristics. Looking at the column for diode 1N4148 we see that the static forward voltage tested with a forward current of 10 mA does not exceed a maximum of 1 V. This specification is also shown on the curve and in effect says that the manufacturer guarantees that the curve of every 1N4148 will pass to the left of this point.

$\mathbf{I_R}$ (reverse current) is also found in the electrical characteristics table. Here we have two test conditions. The first one says that with a reverse voltage of 20 V and air temperature of 25°C the reverse current does not exceed 25 nA (nanoamps). This is another point on the curve. In effect the table guarantees that the curve of each 1N4148 will pass above this point. The second test condition of I_R is at 20 V and 150°C. Here it can be seen how the reverse leakage current increases with temperature. At 150°C the maximum leakage reaches 50 μA (microamps). You will not find this point on the curve as it only covers the 25°C test conditions.

$\mathbf{V_{BR}}$ (reverse breakdown voltage) is shown in the same table. Two limits are given. The first corresponds to the **working peak reverse voltage.** This is the maximum safe reverse voltage that can be applied to the diode before the curve bends sharply and the device breaks down. In this case V_{BR} is 75 V.

The second value of V_{BR} quoted is actually a point on the curve where breakdown is occurring, but where the burnout point has not been reached. Reverse breakdown can be considered to be non-destructive up to this point.

$\mathbf{T_{RR}}$ (reverse recovery time) is found in the switching characteristics table on the second page. The T_{RR} for this diode is 4 ns (4 nanoseconds). This is the time which elapses after forward conduction has stopped, before the diode can start blocking reverse current. It should be noted that the test conditions shown here are fairly typical operating conditions.

Equipped with these five important diode specifications you are now in a position to understand any diode data sheet and to select diodes for use in simple circuits.

Zener diodes

In his research, Zener discovered a very useful characteristic of diodes. He found that altering the level of doping in a diode had an effect on its resistance and its reverse current behaviour.

The I-V curve for a **Zener diode** (one with an altered doping level) is shown in *figure 8*. You can see that in the bottom left hand quadrant the curve remains flat for a certain current level, although the applied voltage varies considerably. This is the normal reverse saturation current of any diode. Beyond a certain reverse voltage the current suddenly rises again, as in *figure 5*. However, in a Zener diode this sudden rise may occur at a much lower voltage (called the Zener voltage). Zener diodes may be designed with Zener voltages from 3 V right up to 150 V. The variation of voltage is achieved by variation of the doping levels.

When arranged in a suitable circuit, a Zener diode can be used as a **voltage regulator** providing regulated, i.e. constant voltages. This is because the voltage across a Zener diode remains

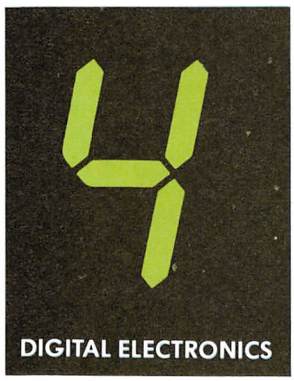
constant as long as the applied reverse voltage is greater than that at which Zener breakdown occurs. When using a Zener diode a series resistance must be included to limit the current flowing so that the Zener diode does not burn out. This series resistor ensures that the voltage across the diode remains constant at the Zener breakdown voltage.

The operating or **dynamic resistance** R_Z is another important characteristic of the Zener diode which can also be seen from the curve. On the horizontal part of the graph the resistance is high, but as the curve enters the knee, resistance is lowered in proportion to the current increase. Ideally, R_Z would be equal to zero since the aim is to produce a constant voltage in the breakdown region – that is to say the curve should be vertical. In reality, R_Z can reach hundreds of ohms. The nearer a Zener diode comes to its ideal characteristics, the better it will be since this means a low dynamic resistance.

Manufacturers usually provide the Zener voltage levels and the maximum current or the maximum power level in their specifications. As before, these amounts are given for a certain specified test temperature: this is because temperature variation can alter the performance of these devices.

Glossary

conventional current	the defined convention for current flow. Current flows from positive to negative in a circuit. By historical accident this is opposite to the direction of electron flow
forward current (I_F)	the amount of current flowing in a diode when its anode is at a more positive voltage than its cathode
forward voltage (V_F)	the voltage present across a diode when its anode is at a more positive voltage than its cathode
power dissipation rating	the amount of power a diode can handle, dissipating the excess energy in the form of heat
reverse breakdown voltage (V_{BR})	the point at which a diode ceases to block a reverse voltage and breaks down
reverse current (I_R)	the amount of current which flows through a diode when a reverse voltage is applied
reverse recovery time (T_{RR})	the time taken for a diode to conduct after it has been blocking a voltage. This time is critical when high frequency alternating currents are being rectified
reverse voltage (V_R)	the voltage across a diode when its cathode is at a more positive voltage than its anode
Zener breakdown voltage	the voltage at which electrons are separated from the covalent bond of a reverse biased diode
Zener diode	a particular type of diode which takes advantage of the Zener breakdown voltage to maintain a constant voltage across itself
Zener voltage	the constant voltage present across a Zener diode when it is working in reverse current mode



The TTL logic family

History and development

Along with MOS devices, the TTL family is one of the most widely used in digital electronics. TTL stands for **Transistor Transistor Logic.** In this chapter we will look at the characteristics and operation of a number of TTL ICs and the way they are connected in circuits.

As recently as twenty years ago, electronic equipment used discrete components: resistors, capacitors and valves. This equipment was very bulky, had high energy consumption and dissipated a lot of heat – this meant that early digital machinery was not very reliable or economic. Transistors, developed in 1948, are small, use little energy and are mass produced very cheaply. Because of their reliability they quickly replaced valves.

The basic idea of integrated circuits was formulated in the late fifties and early sixties. By 1965 ICs containing up to thirty transistors on a single silicon chip were being produced. The first ICs were still fairly big, but more and more development continued to reduce the size of these devices and increase the level of integration possible.

Around 1970, small scale integration (SSI) gave way to medium scale integration (MSI) which increased the number of gates per chip from 10 to 100. Large scale integration (LSI) with over 100 gates per chip followed. Very large scale integration (VLSI) components are now commonplace. These VLSI devices have more than 1000 gates per chip and are widely used to make complete microprocessor ICs. Even greater levels of integration are now available for specialized devices.

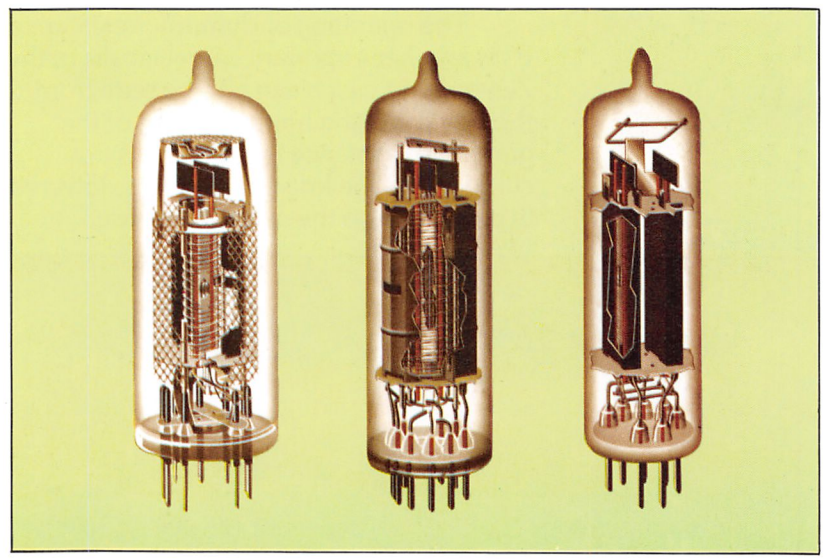
Integrated components

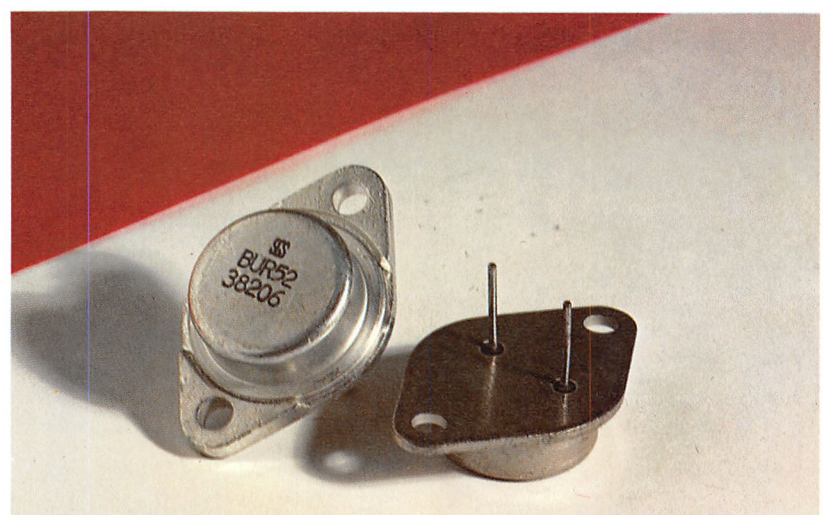
During the production of ICs, hundreds of chips are manufactured on a single slice of silicon. These are electrically tested and then separated.

The chips that pass this test are assembled into special containers called **packages** which are then finally tested.

These packages differ according to the material used and the number of connection pins on each one. The number of pins can vary from eight to as many as forty, as used in microprocessors. The most common ICs have 14 pins.

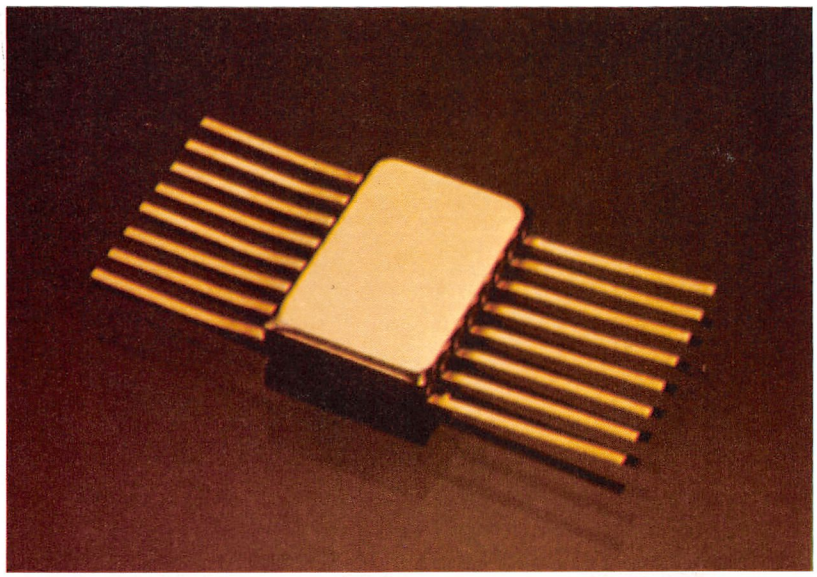
Packages can be divided into two





Top: cut-away diagram of three valves of the type used in first generation computers. (Philips).

Two power transistors, actual size. (SGS).



Above: Integrated circuit in an F-type (flat) ceramic package. (Mostek)

other categories according to their layout. One type is **flat** and has the commercial mark F; the other is **dual in line** (DIL), so called because it has two parallel rows of connecting pins. DIL devices are marked N if the package is plastic and J if it is ceramic. There is also a special TO-5 type which is used for linear integrated circuits.

ICs are more correctly called **integrated electronic components** because the chip at the heart of the component usually carries more than one actual working circuit. The simple TTL ICs covered in this chapter can contain up to 6 independent gates per 14 pin package, each gate being made of a single (integrated) circuit. These gates exist side by side on the chip. Other ICs are available which perform more complex functions and these contain thousands of circuits.

These integrated electronic components are known as **monolithic** integrated circuits, because all the individual circuits exist on one chip (mono = one). Other devices, called **hybrid** integrated circuits, are made up of several separate chips interconnected within the same package.

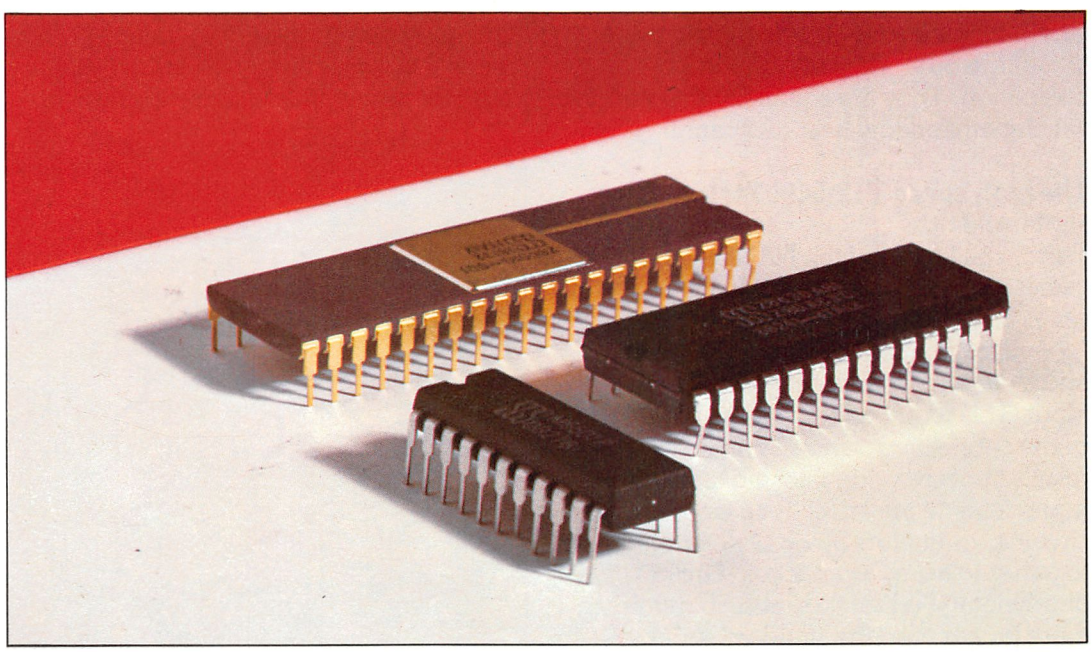
The standard 54/74 series

The first TTL devices, the 54 series, were marketed at the end of the sixties and were first used in military applications. Later on, as prices dropped, they developed into the 74 series which were used commercially.

As new production techniques were developed, other TTL subfamilies were produced. For example, the standard SN/54 series, the fast SN54H/74H series, the reduced power absorption series, SN54L/74L, and the Schottky SN54S/74S series. The code letters SN, first introduced by Texas Instruments, mean **Semiconductor Network.**

Although each of these families was originally developed for different applications, they are all directly compatible and can be interfaced with one another.

Right: Integrated circuits in Dual in Line (DIL) packages. (SGS)



Typical characteristics and subfamilies

Typical characteristics of TTL devices:

Supply	5.0 V
Logic 0 level	0.2 V
Logic 1 level	3.0 V
Noise immunity	1.0 V

Before we examine them in detail let's look at the complete list of standard products in the 54/74 series. Although Texas Instruments' devices have been chosen, they are basically the same as any other manufacturer's components. The list has been divided according to the degree of integration of the devices. *Table 1* shows MSI devices which are monolithic circuits performing more complex functions than SSI devices.

The 54L/74L low power absorption family

Figure 1 shows the basic circuit of a NAND low power absorption gate. It differs from the standard 54/74 series (see *Digital Electronics – 3, figure 8*) in that it uses higher value resistors. Since, for a fixed supply voltage, an increase in resistance corresponds to a reduction in the power dissipated, the power required is about one-tenth of that needed by other TTL ICs.

Devices in the 54L/74L series present a power dissipation of about 1 mW per gate. Their speed is approximately twice that of the other circuits with the same dissipation. They are useful for applications which demand low heat dissipation.

The high speed TTL 54H/74H subfamily

The circuit layout of the NAND 54H/74H gate is also similar to the standard 54/74 series and is shown in *figure 2*.

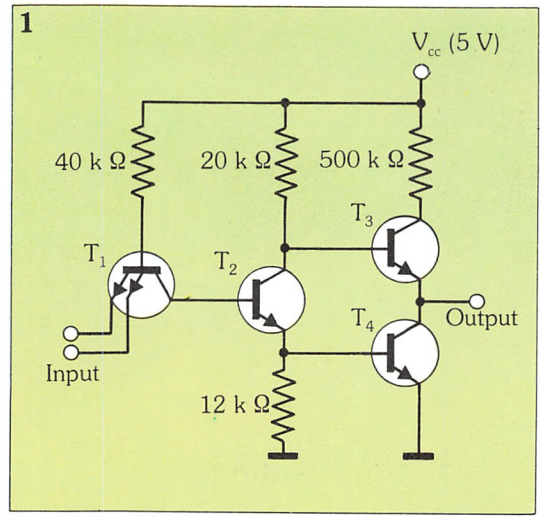
Here, the resistance values are low and a diode is inserted on every emitter input. This reduces the capacitive effects of the connecting transmission lines, which results in decreased switching times. This 'transmission line effect' is caused by the circuit's conductors being close enough together to act as a capacitor. Their disadvantage is that they absorb much more power than the standard series.

Table 1. Standard TTL integrated circuits

Medium-Scale Integration (MSI)

Decoders:	
BCD-to-decimal (4-to-10 lines)	SN54/7442
Excess-3-to-decimal (4-to-10 lines)	54/7443
Excess-3-Gray -to-decimal (4-to-10 lines)	54/7444
BCD-to-decimal decoder/driver (4-to-10 lines)	54/7445
BCD-to-7 segment decoder/driver	54/7446
BCD-to-7 segment decoder/driver	54/7447
BCD-to-7 segment decoder/driver	54/7448
BCD-to-7 segment decoder/driver	54/7449
BCD-to-decimal decoder/driver	SN74141
BCD-to-decimal decoder/driver	SN54/74145
4-line-to-16-line decoder/demultiplexer	54/74154
Dual 2-line-to-4 line decoder/demultiplexer	54/74155
Dual 2-line-to-4 line decoder/demultiplexer, open collector	54/74156
Memories/latches:	
4-bit bistable latches	SN54/7475
4-bit bistable latches	54/7477
16-bit active element memory	54/7481
16-bit active element memory	54/7484
256-bit read-only memory	SN7488
8-bit bistable latches	SN54/74100
Arithmetic elements:	
Gated full-adder	SN54/7480
2-bit full-adder	54/7482
4-bit full-adder	54/7483
Quad 2-input exclusive-OR gate	54/7486
4-bit arithmetic logic unit	54/74181
Look-ahead carry generator	54/74182
Counters:	
Decade	SN54/7490
Divide-by-12	54/7492
4-bit binary	54/7493
Synchronous decade up/down	54/74192
Synchronous 4-bit up/down	54/74193
Shift registers:	
8-bit	SN54/7491A
4-bit	54/7494
4-bit	54/7495
5-bit	54/7496
Data selectors/multiplexers:	
16-bit	SN54/74150
8-bit, with strobe	54/74151
8-bit	54/74152
Dual 4-line-to-1-line	54/74153
Miscellaneous:	
8-bit odd/even parity generator/checker	SN54/74180

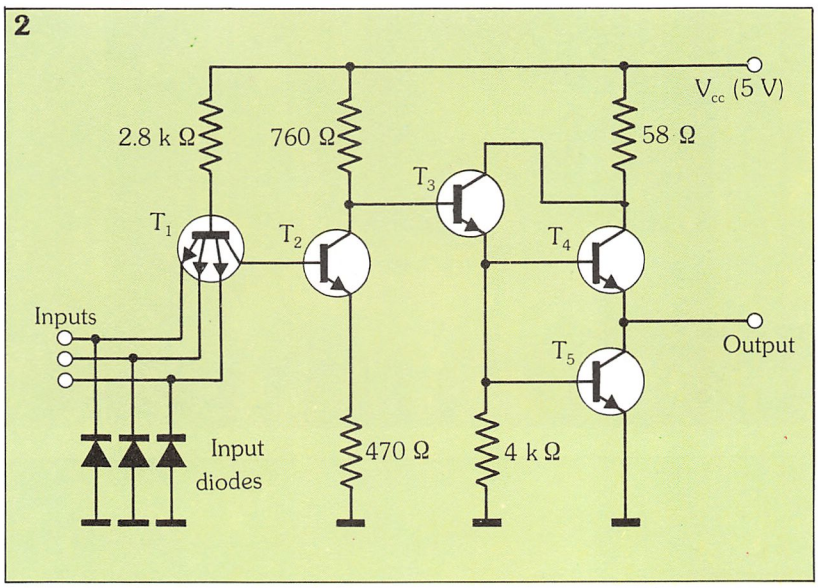
The output section employs a Darlington pair (transistors T_3 and T_4). This means they are arranged to act as a single transistor with a greater current gain

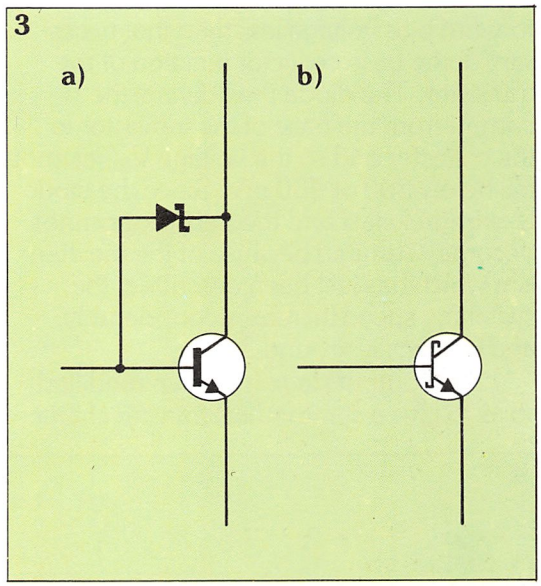


1. Basic circuit layout of an SN54L/74L NAND gate.

(the gain is, in fact, the product of the two individual transistors' gains). The Darlington pair also helps to reduce switching times–a typical time being about 6 ns per gate.

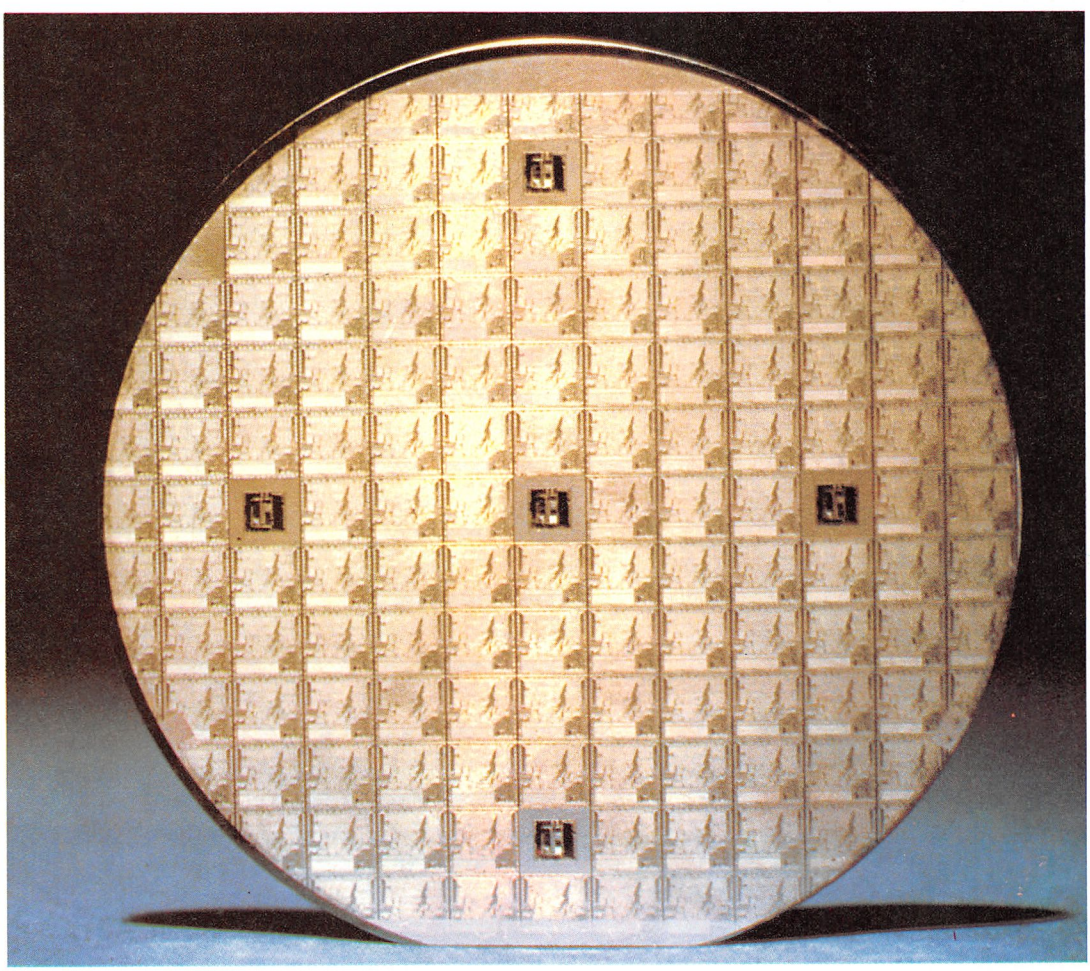
The Schottky TTL 54S/74S subfamily
The TTL Schottky series is the fastest. It combines the high speed of emitter coupled logic (ECL) with the relatively low TTL power consumption. This takes place





2. Basic circuit layout of an SN54H/74H NAND gate.

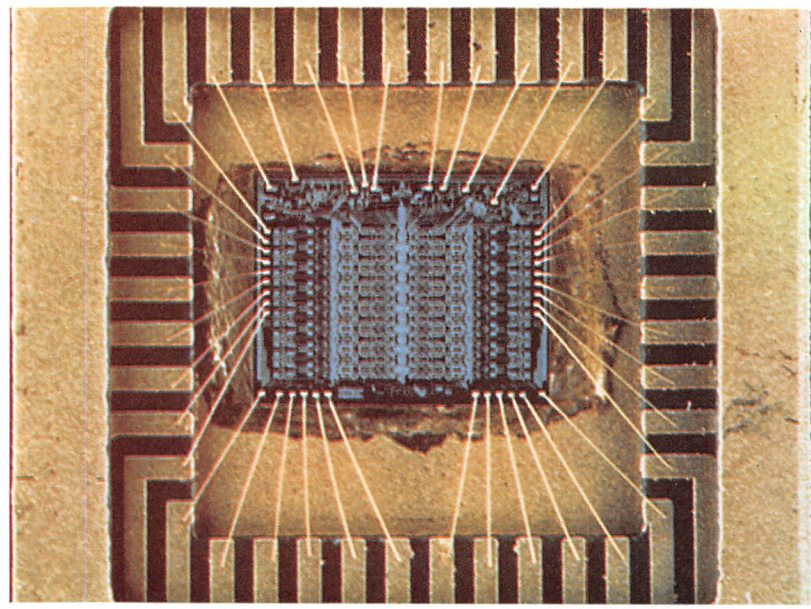
3. Basic circuit arrangement, showing a) transistor with Schottky barrier diode and b) symbol for transistor with built-in Schottky diode.

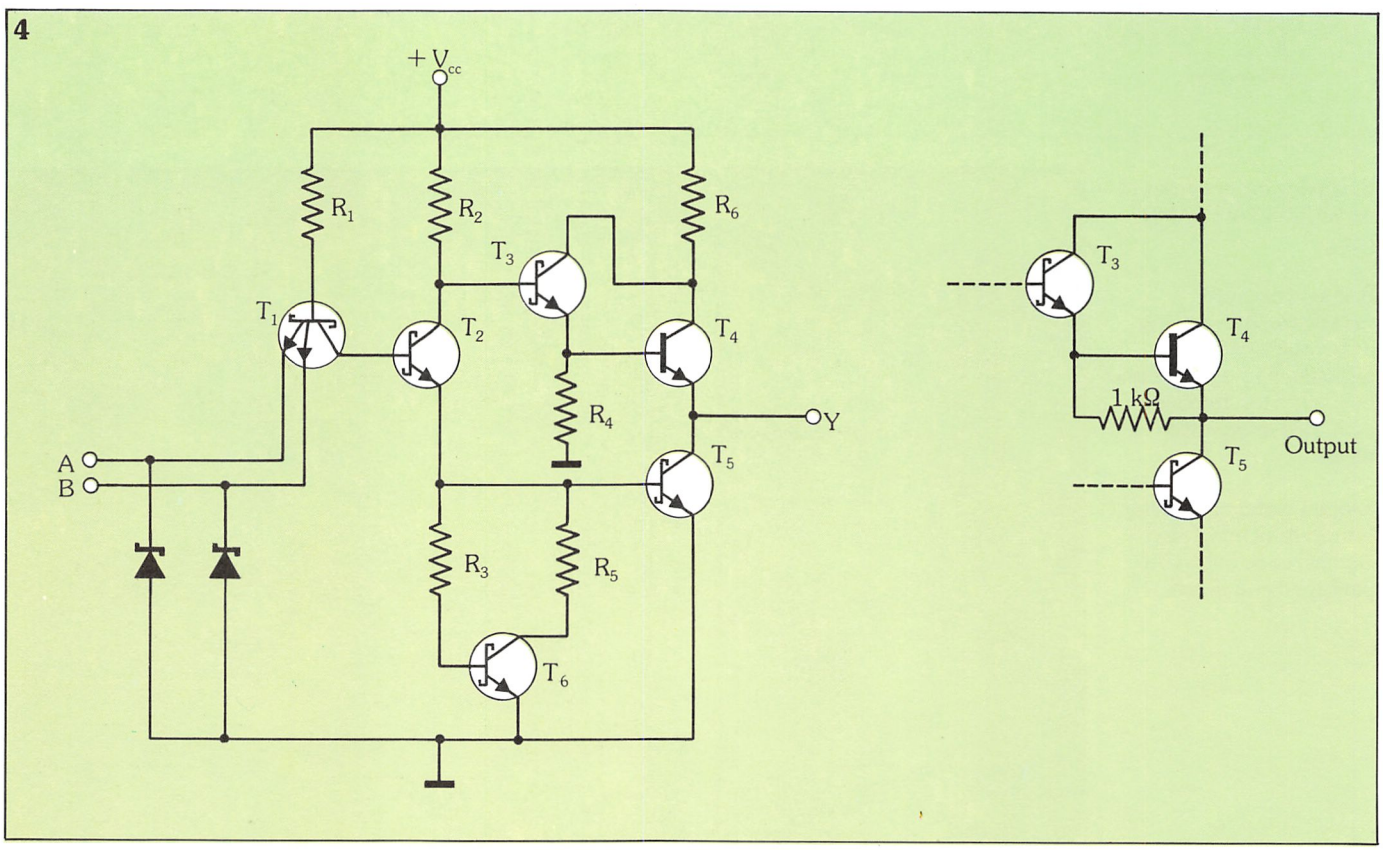


Silicon wafer containing hundreds of integrated circuits ready to be cut, packaged and tested.

through the use of a Schottky barrier diode as a blocking or clamping diode between the base and collector (*figure 3*). The particular characteristics of a Schottky diode prevent the transistor from becoming saturated or 'bottomed'. It has a forward bias voltage less than that necessary at the base collector junction of the transistor. The diode then diverts the current from the base of the transistor to the collector and so the voltage V_{BC} cannot fall below 300 or 400 mV. Since the diode is extremely fast and the transistor cannot become saturated because of the mechanism which has just been described, the switching speed increases considerably, and is typically around 3 ns.

The output stage has been modified so as to give a symmetrical transfer charac-





teristic. In *figure 4*, transistor T_6 and the associated resistors take the place of the resistance between the base of the final transistor (T_5) and earth. The power dissipated per gate is about 20 mW.

In recent years, a circuit arrangement known as 54LS/74LS has become more and more widespread. The term LS means low power Schottky TTL. It is basically the same as the previous Schottky series except the base resistance of transistor T_4 is connected directly to the output instead of to earth (on the right in *figure 4*). The power dissipated is much reduced although the propagation delay is increased.

Integrated circuit, with the wires already bonded, ready for packaging.

4. Circuit design for an 54/74S NAND gate. The variation on the right is for the 54/74LS NAND gate.

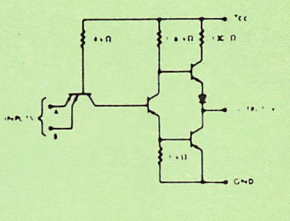
Specifications

Before introducing the TTL gates which perform the various logic functions, it might be useful to say a few words about data sheets. The data sheet is supplied by the manufacturer and contains the specifications of a device, together with the information necessary for the correct operation of the component. An example for the Texas Instruments SN 7400 IC is shown in *figure 5*.

5. Data sheet for the Texas Instruments SN 7400 IC.

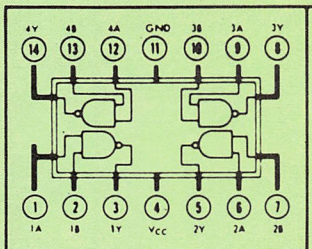
CIRCUIT TYPES SN5400, SN7400
QUADRUPLE 2-INPUT POSITIVE NAND GATES

schematic (each gate)

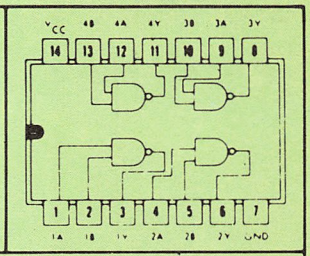


NOTE: Component values shown are nominal.

W FLAT PACKAGE (TOP VIEW)



J OR N DUAL-IN-LINE PACKAGE (TOP VIEW)



positive logic: $Y = \overline{AB}$

recommended operating conditions

	MIN	NOM	MAX	UNIT
Supply Voltage V_{CC}: SN5400 Circuits	4.5	5	5.5	V
SN7400 Circuits	4.75	5	5.25	V
Normalized Fan-Out From Each Output, N			10	
Operating Free-Air Temperature Range, T_A: SN5400 Circuits	−55	25	125	°C
SN7400 Circuits	0	25	70	°C

electrical characteristics over recommended operating free-air temperature (unless otherwise noted)

PARAMETER		TEST FIGURE	TEST CONDITIONS†		MIN	TYP‡	MAX	UNIT
$V_{in(1)}$	Logical 1 input voltage required at both input terminals to ensure logical 0 level at output	1			2			V
$V_{in(0)}$	Logical 0 input voltage required at either input terminal to ensure logical 1 level at output	2					0.8	V
$V_{out(1)}$	Logical 1 output voltage	2	V_{CC} = MIN, V_{in} = 0.8 V, I_{load} = −400 μA		2.4	3.3		V
$V_{out(0)}$	Logical 0 output voltage	1	V_{CC} = MIN, V_{in} = 2 V, I_{sink} = 16 mA			0.22	0.4	V
$I_{in(0)}$	Logical 0 level input current (each input)	3	V_{CC} = MAX, V_{in} = 0.4 V				−1.6	mA
$I_{in(1)}$	Logical 1 level input current (each input)	4	V_{CC} = MAX, V_{in} = 2.4 V				40	μA
			V_{CC} = MAX, V_{in} = 5.5 V				1	mA
I_{OS}	Short-circuit output current§	5	V_{CC} = MAX	SN5400	20		−55	mA
				SN7400	−18		−55	
$I_{CC(0)}$	Logical 0 level supply current	6	V_{CC} = MAX, V_{in} = 5 V			12	22	mA
$I_{CC(1)}$	Logical 1 level supply current	6	V_{CC} = MAX, V_{in} = 0			4	8	mA

switching characteristics, V_{CC} = 5 V, T_A = 25°C, N = 10

PARAMETER		TEST FIGURE	TEST CONDITIONS	MIN	TYP	MAX	UNIT
t_{pd0}	Propagation delay time to logical 0 level	65	C_L = 15 pF, R_L = 400 Ω		7	15	ns
t_{pd1}	Propagation delay time to logical 1 level	65	C_L = 15 pF, R_L = 400 Ω		11	22	ns

† For conditions shown as MIN or MAX, use the appropriate value specified under recommended operating conditions for the applicable device type.

‡ All typical values are at V_{CC} = 5 V, T_A = 25°C.

§ Not more than one output should be shorted at a time.

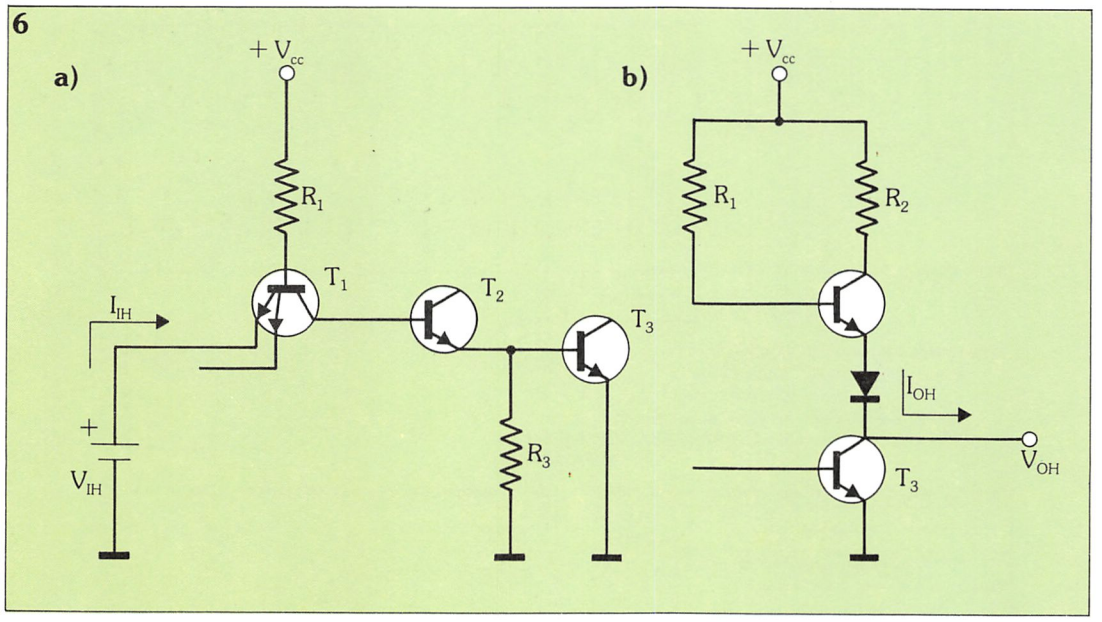
Immediately after the code of the integrated circuit (in this case SN 5400 and SN 7400), the logical function performed is given. This is a device which contains four positive NAND gates, with two inputs each (quadruple 2 input positive NAND gates).

After the electrical circuit relating to each gate, the specifications for each type of package available are given. The component block diagram shows the numbering of the pins and their electrical connections. In this case, for the DIL package to perform a positive NAND function, the input levels must be introduced on pins marked 1 and 2, 4 and 5, 9 and 10, 12 and 13. The output logic levels will be taken from pins 3, 6, 8 and 11. Should only one gate of the package be used, only two inputs and one output will be chosen. The pins marked 7 and 14 are the earth and supply connections.

The next item on the data sheet gives us the recommended operating conditions. Here the manufacturer gives the rated voltage, the minimum and maximum supply voltages, the maximum fan-out for each output and the permitted temperature variation.

The large table in *figure 5* refers to the electrical characteristics. The first parameter shown is V_{in} (1). It is the threshold voltage for level H which means high level input voltage. V_{IH} (voltage input high) is another name for this. The integrated circuit in this case represents a NAND gate. So, to get logic level 0 on the output, a 1 signal must be applied to both inputs. This condition is guaranteed when the input voltage is at least 2 V.

V_{in} (0) or V_{IL} is the logic level 0 required at both inputs to ensure a logic 1 output. It represents the threshold voltage for level L and must not exceed 0.8 V. V_{out} (1) or V_{OH} is the output voltage for level H. It must have a minimum of 2.4 V while a typical value is 3.3 V. V_{out} (0) or V_{OL} is the output voltage with level 1 on the inputs. Its maximum is 0.4 V and its typical level is 0.22 V.



6. Partial diagram of an SN5400 electrical circuit showing the flow of current with a) input high and b) output high.

7. Partial diagram of an electrical circuit with a) input low and b) output low.

8, 9 and 10. Examples of the output characteristics of some gates.

I_{in} (0) or I_{IL} is the input current for level L. It represents the current which flows in an input when a low voltage is applied. This is determined by applying the maximum permitted supply voltage and level 0 of 0.4 V. The maximum allowable value is −1.6 mA (the minus sign appears when the current flows from the device to outside).

I_{in} (1) or I_{IH} is the input current for level H. This represents the current flowing in the input when a high voltage is applied. Note that with maximum supply voltage (V_{cc} = max), I_{IH} is equal to 40µA or 1 mA, according to whether input V_{in} is equal to 2.4 V or 5.5 V.

I_{OH} and I_{OL} are the output currents for level H (high) and L (low). I_{CC} (0) or I_{CCL} represents the current which flows in

Right: an MC68000L microprocessor. (Photo: courtesy Motorola).

pin 14 when all the gates give output 0. t_{PD} (0) or t_{PHL} is the propagation delay time from H to L. It is typically 7 ns when:

$V_{CC} = 5$ V, $R_L = 400\ \Omega$ and
$C_L = 15$ pF.

Finally consider the propagation delay time from L to H: t_{PD} (1) or t_{PLH}. Under test conditions it is 11 ns. If the rated supply voltage is 5 V, then the fan-out (as explained in the last chapter) for level H can be calculated from the maximum input and output currents given:

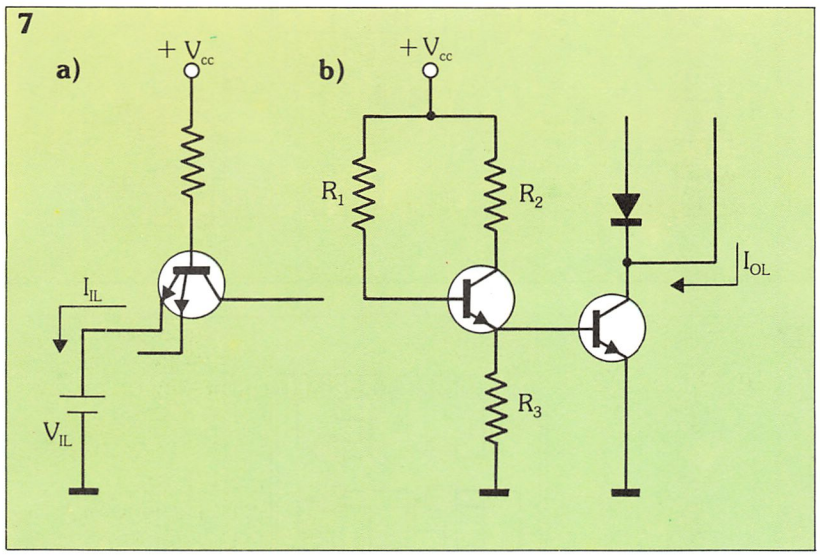
$$\frac{I_{OH}\ max}{I_{IH}\ max} = \frac{400\ \mu A}{40\ \mu A} = 10$$

and for level L:

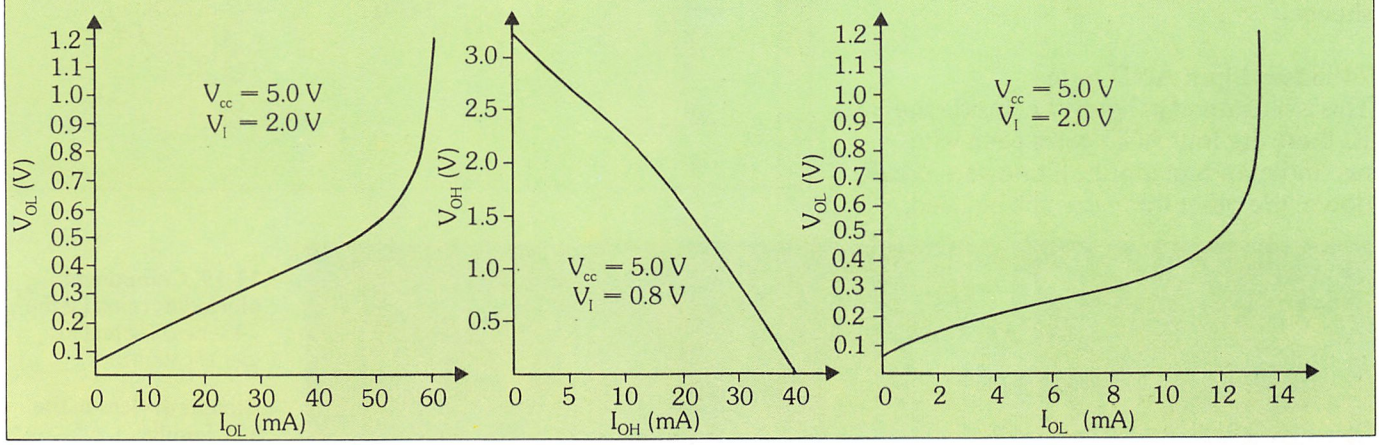
$$\frac{I_{OL}\ max}{I_{IL}\ max} = \frac{16\ \mu A}{1.6\ \mu A} = 10$$

When an input is at level H, the input current I_{IH} is entering and therefore positive. This can be seen in *figure 6a*. If, instead, the output is at level H (*see figure 6b*), the current I_{OH} comes out of the terminal and so is negative. Vice versa, if the input is at level L, the I_{IL} is negative and if the output is at level L, the I_{OL} current becomes positive. This can be seen in *figure 7a* and *b*. Since the current I_{OH} is supplied by the driver logic element, it is called **source current.** The I_{OL} current, known as **sink current,** comes from the load instead and is absorbed by the logic element.

7

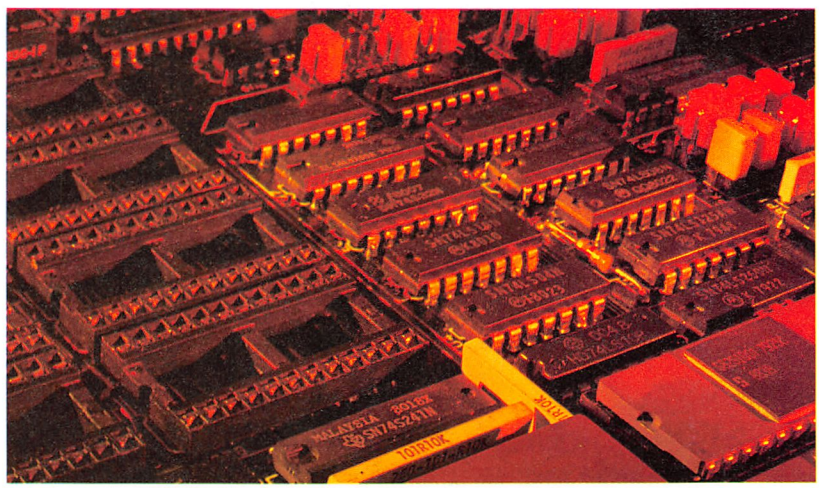


8



9

10



Input and output characteristics

The input characteristic graphically represents the behaviour of the input voltage V_I in relation to the variation of the current I_I. The output characteristic shows the variation of V_O as a function of the current I_O. *Figures 8* to *10* show some examples.

The 74 series ICs

Let's now look in detail at the integrated devices in the 74 series. We will limit this to the basic AND, OR, NOT, NAND and NOR gates that have been covered so far. All the gates in this series are made with 14 pins on the chip. Two of these pins are for earth and positive supply, leaving twelve for the input and output gates.

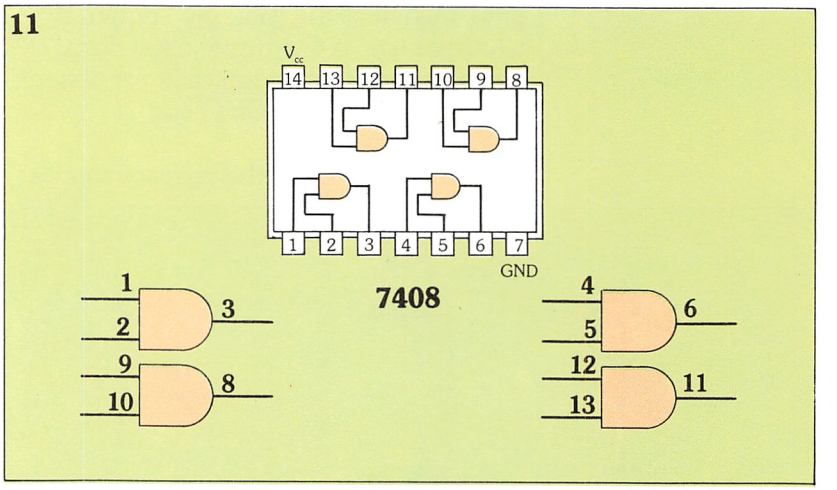
Various input/output combinations are possible:

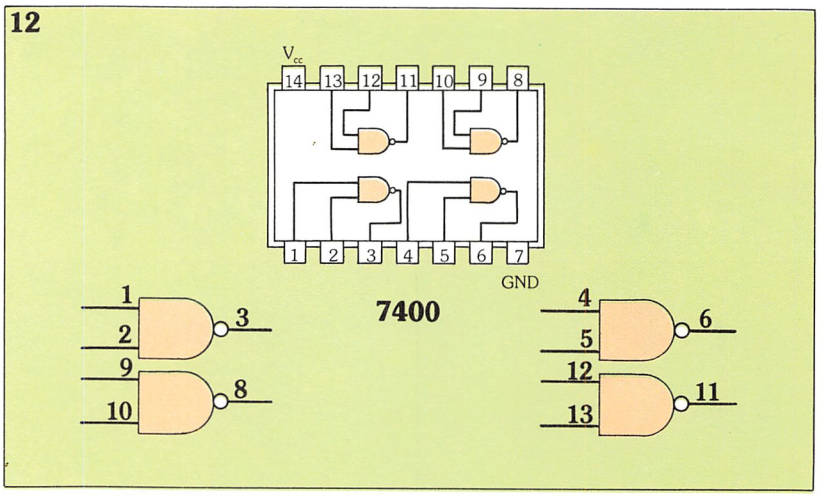
–six gates: each one input one output;
–four gates: each two inputs one output;
–three gates: each three inputs one output;
–two gates: each five inputs one output;
–one gate: up to eleven inputs one output.

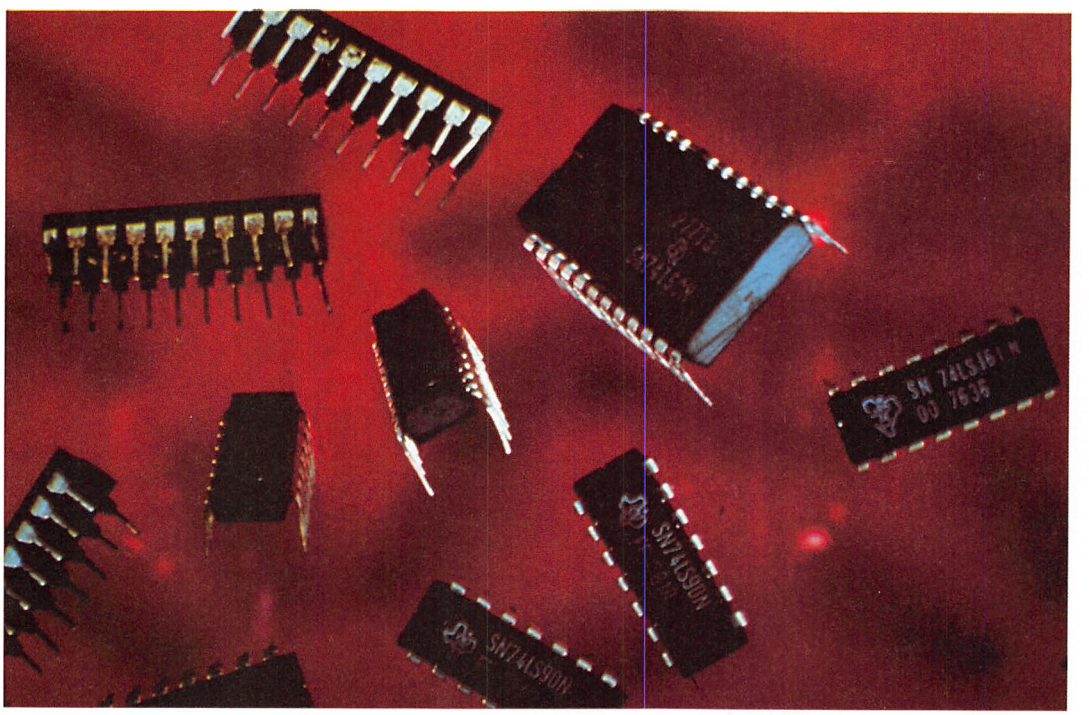
Figures 11-19 show examples of 74 series ICs, featuring the component block diagrams that would appear on the data sheets.

7408 two input AND gate

This is illustrated in *figure 11*. Inside the IC there are four AND gates each with two independent inputs. If one gate breaks down, the other three can still be used.



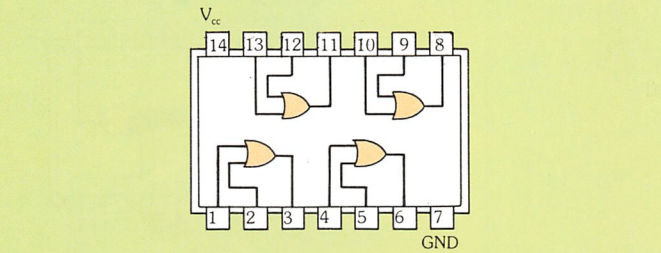




11-19. Component block diagrams for the 74 series (see text for identification).

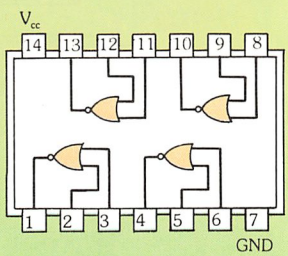
Bipolar devices in the TTL family.

13

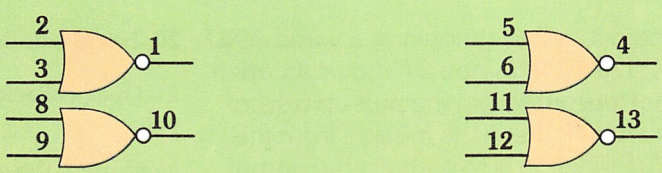


7432

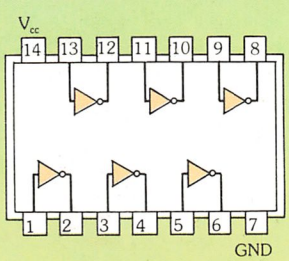
14



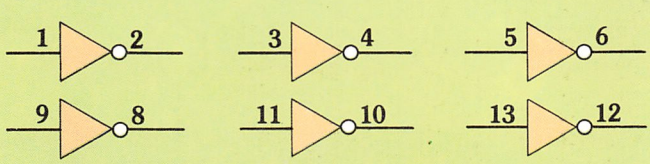
7402



15



7404



7400 two input NAND gate
The configuration of the pins of this chip 7400 two input NAND gate (*figure 12*), is similar to that of the 7408 IC. The supply pin is 14. As in the 7408, pin 7 is the earth (ground-GND) and the chip contains four NAND gates each with two independent inputs.

7432 two input OR gate
Figure 13 shows the configuration of the pins of a 7432, two input OR gate. It is similar to that of the 7400 and 7408 chips. Note that in all these three chips the input and output pins correspond. In chip 7432 there are four independent gates.

7402 two input NOR gate
Figure 14 shows the configuration of the pins of the 7402 two input NOR gate. It is made up of 4 independent NOR gates. It can be seen that the layout of the gates is quite different from those of the preceding integrated circuits in the range.

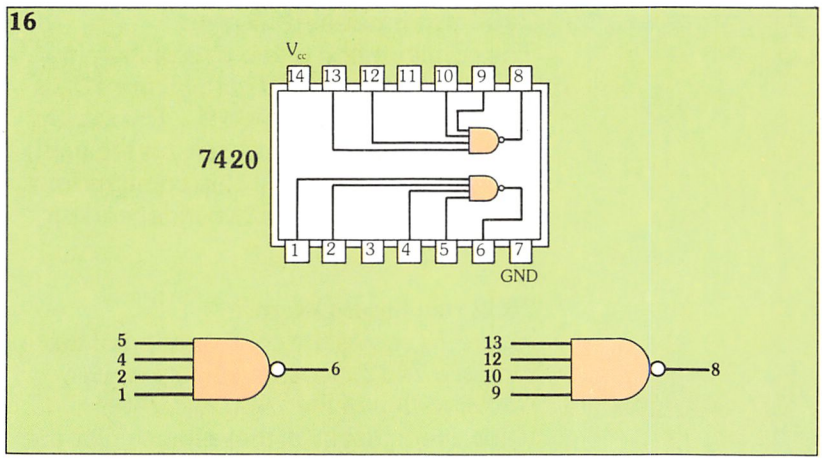
7404 inverter
The 7404 inverter chip contains 6 independent inverters (NOT gates) as shown in *figure 15*.

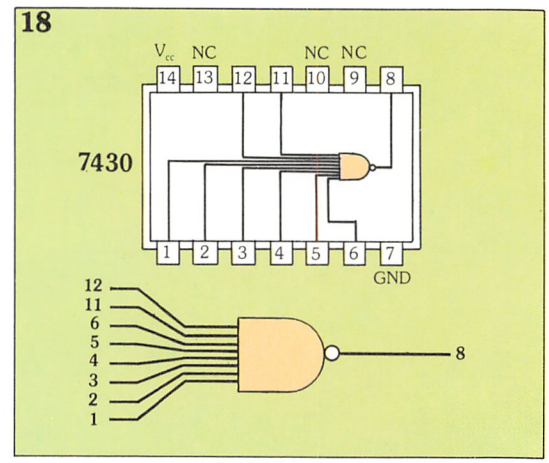
7420 four input NAND gate
The 7420 double NAND gate chip contains only two independent gates each with four inputs (see *figure 16*).

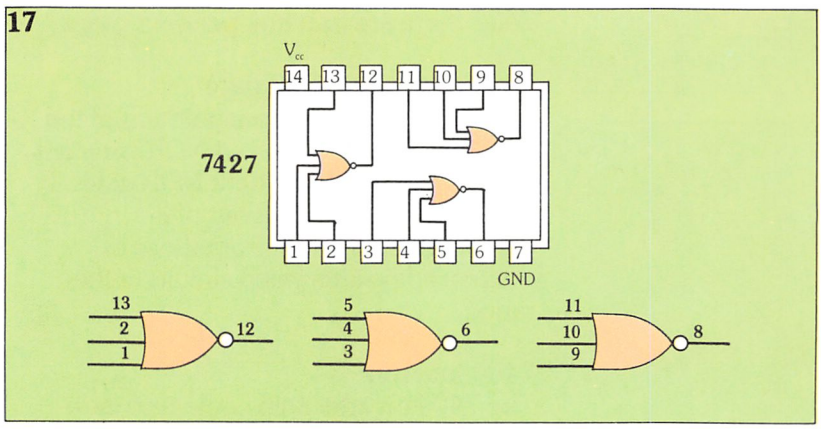
7427 three input NOR gate
Figure 17 shows the internal arrangement of the 7427 chip which contains 3 NOR gates each with three independent inputs.

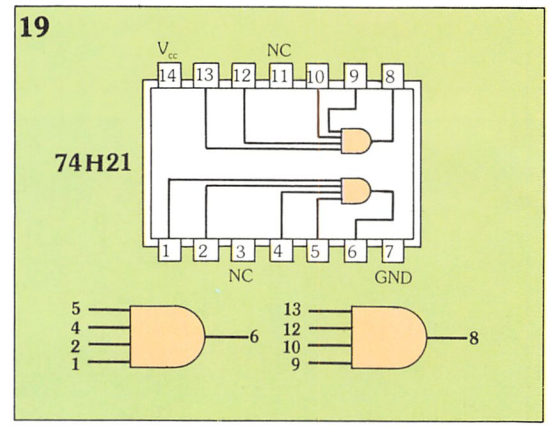
7430 eight input NAND gate
The 7430 illustrated in *figure 18*, contains a single NAND gate, having eight independent inputs.

74H21 four input AND gate
The 7421 is a double four input AND gate. As can be seen from *figure 19*, it is similar to the 7420 in the configuration of the pins. It is a high speed chip (as shown by the H in its number) and can operate at a higher switching speed than the normal 7421 chip.









Totem pole and open collector outputs TTL integrated circuits have two different output configurations: **totem pole** and **open collector.** A brief explanation of what these mean will do for the moment, because they will be examined in detail in a later chapter.

Go to the TTL gate shown in *figure 20.* Look a bit more closely at the final stage transistors T_3 and T_4. When the output is at logic level 1, T_4 behaves as an emitter follower which drives current into the load.

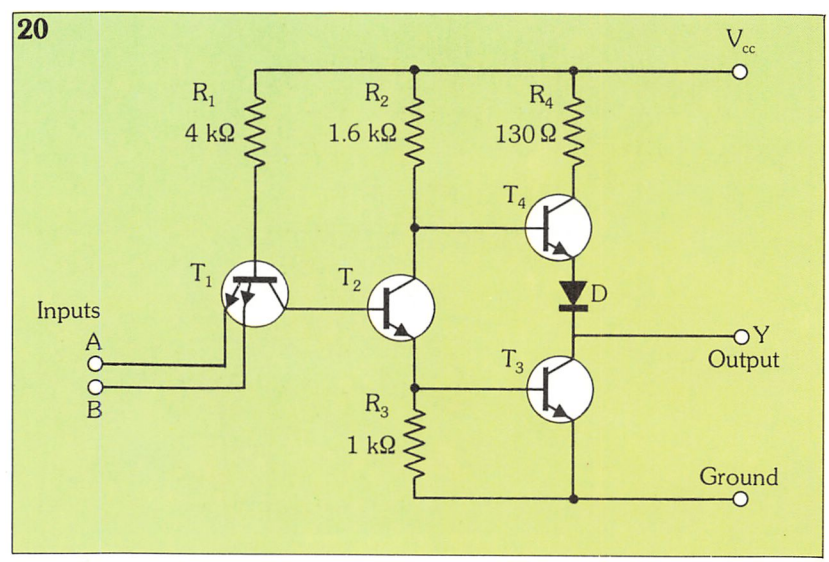
If instead, it is at logic 0, current coming from the load 'sees' only the bottoming resistance of T_3 which is very low. So the configuration presents a low output impedance in both states. This allows capacitive loads to be driven, and ensures the high speed performance of the TTLs. This is known as a **totem pole output.**

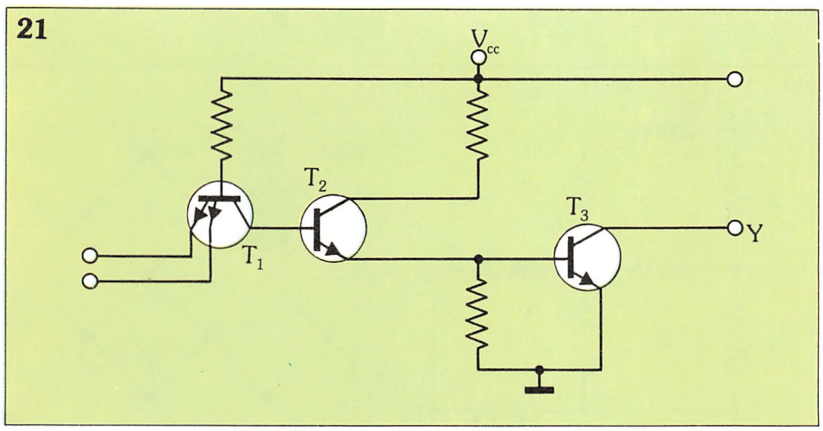
The disadvantage of the totem pole output is the difficulty found when two or more outputs of separate gates are to be connected together (known as a wired AND or WIRED-AND). *Figure 21* shows an **open collector output** stage; the pull-up resistor (R_4), seen in *figure 20,* is missing from this circuit. Its function is to connect the output to the high level. When several outputs are joined together one pull-up resistor can be

20. Totem pole output.

21. Open collector output.

22,23. Wired AND connections.

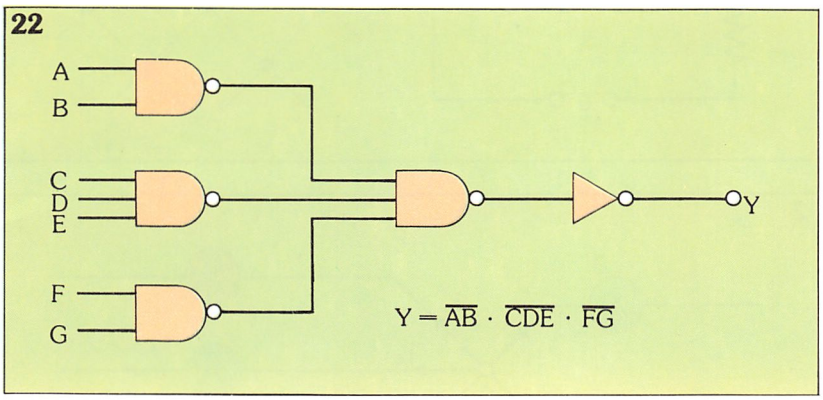


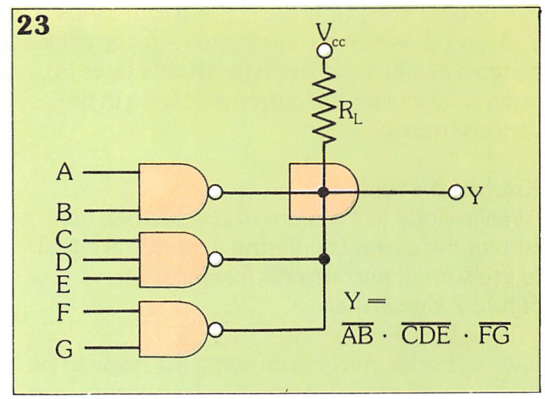


used (R_L in *figure 23*), and its resistance depends on the number of outputs which are to be connected together.

To make a wired AND configuration with totem pole outputs the scheme shown in *figure 22* would have to be used, but this uses more ICs.

This wired OR connection is useful in specialized circuits but the totem pole output is more common. Some ICs are designed for exclusive OR functions which include two OR gates with the collectors wired together (as above) but all fabricated on a single chip.





Glossary

data sheet	the technical specification and connection details for an individual IC
DIL	dual in line. IC package which has two parallel lines of connecting pins. The pins are 0.1 inch apart and can 'plug' into a circuit board
LSI	large scale integration. A generation of ICs with more than one hundred gates on each chip, to a maximum of a thousand
MSI	medium scale integration. A generation of ICs with a maximum of one hundred gates on each chip
open collector	TTL output stage without an inbuilt pull-up resistor. This enables many circuits to be joined in a WIRE-AND configuration, using a single resistor
SSI	small scale integration. A generation of ICs with a maximum of ten gates per chip
totem pole	an output configuration of a TTL IC, which presents its load with a low impedance in both the on and off states. This enables it to drive capacitive loads and to preserve its high speed switching capability
VLSI	very large scale integration, a generation of ICs which have over a thousand gates per chip

ELECTRICAL TECHNOLOGY

Methods of circuit analysis

Ohm's law (which was explained in a previous chapter) forms the basis of electrical circuit analysis, however, Ohm's law alone is not adequate for the analysis of the various complex circuits which can occur in practice. An electrical network may be made up of components which are linked in a very complex way in various branches. Each point at which three or more branches converge is known as a node, and a closed circuit formed by three or more consecutive branches is known as a loop.

The electrical network of *figure 1* contains five nodes, A, B, C, D and E, and four loops, 1, 2, 3, and 4. When we are faced with complex systems like this, we use **Kirchhoff's laws** in order to calculate the currents flowing in the various branches.

Kirchhoff's first law

At each node in a system of conductors, the sum of the currents entering the node is equal to the sum of the currents leaving it.
Figure 2 shows that:

$$I_1 + I_2 = I_3 + I_4 + I_5$$

If we consider currents entering the node to be positive and those leaving it to be negative, we find that the algebraic sum of the currents is zero, and in this case the following can be written:

$$I_1 + I_2 - I_3 - I_4 - I_5 = 0$$

Kirchhoff's second law

In each circuit loop, the sum of the products of the current and resistance in each branch is equal to the algebraic sum of the electromotive forces in the circuit.

If we assume clockwise currents to be positive and anticlockwise currents to be negative, the circuit of *figure 3* gives:

$$E_1 - E_3 = (R_1 \cdot I_1) - (R_2 \cdot I_2) - (R_3 \cdot I_3) + (R_4 \cdot I_4)$$

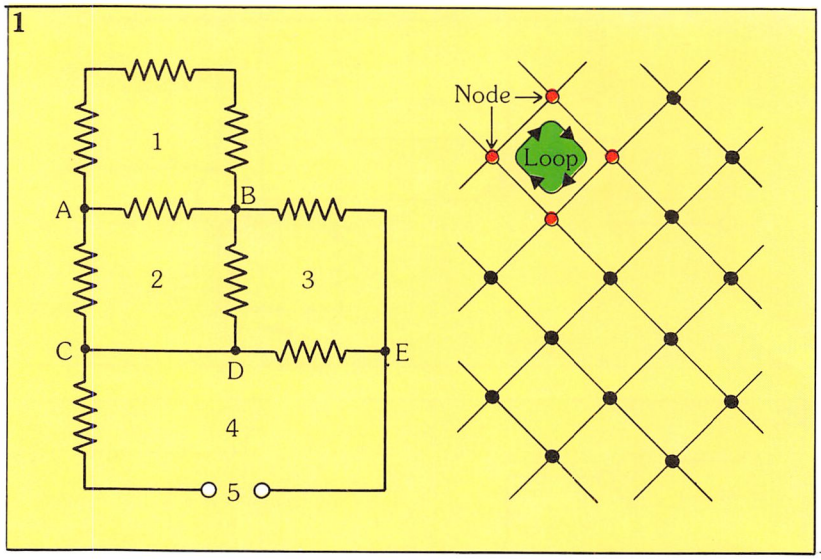
Alternatively this may be written to say that the algebraic sum of the voltages on the various branches of each closed loop is always equal to zero. Thus:

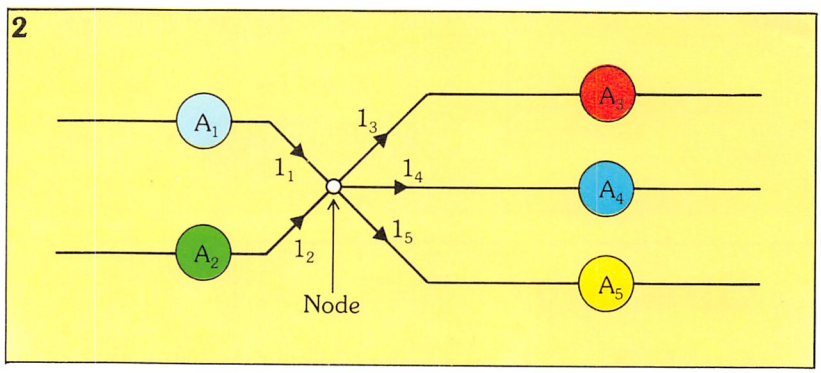
$$V_{AB} + V_{BC} + V_{CD} + V_{DA} = 0$$

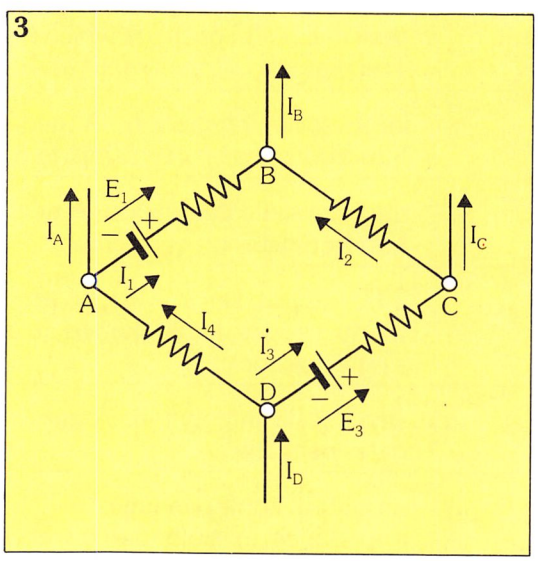
Apart from the two laws, other Kirchoff methods of resolution are very useful. One of these is **Thévenin's theorem.** This permits calculations relating to a branch of a complex network to be simplified by replacing it with an ideal voltage generator and a resistance.

Thévenin's theorem

Thévenin's theorem states that the current in any branch of a network is identical to the current which would flow if the entire network was replaced by an equivalent circuit comprising an ideal voltage generator and a series resistance.



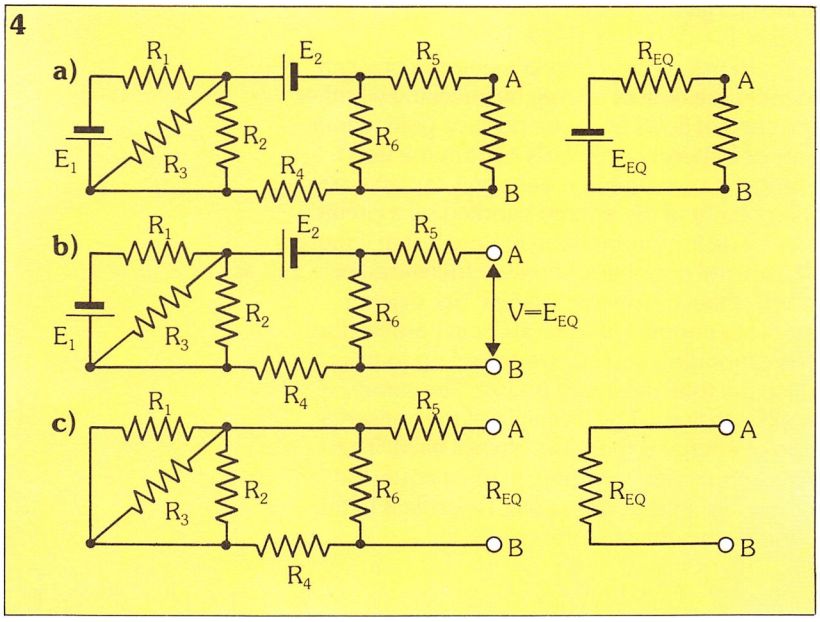




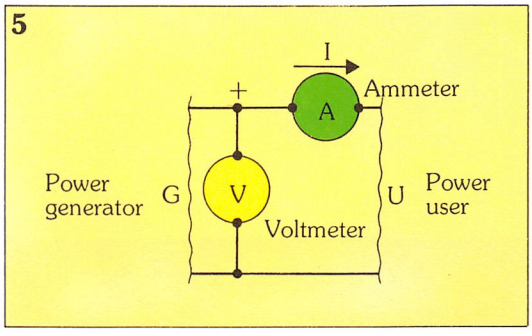
1. An example of an electrical circuit with five nodes and four loops.

2. Electrical analysis of a circuit node, using Kirchhoff's first law.

3. Kirchhoff's second law, used to analyse an electrical loop.



4. Stages in the analysis of an electrical circuit, using Thévenin's theorem.



5. Part of an electrical circuit, used to demonstrate the principle of the conservation of energy.

The EMF of such a generator corresponds to the potential difference which the network produces at the open terminals of the branch concerned. The series resistance corresponds to the equivalent resistance of the network acting as the internal resistance of the generator.

Figure 4 shows a circuit in which we wish to know the current in branch A-B. The equivalent circuit, according to Thévenin's theorem, is as shown at the side. In order to evaluate the EMF of the ideal generator, the voltage drop is calculated between A and B in the circuit of *figure 4b*. Kirchhoff's laws are applied in each of the loops and at the nodes as explained earlier, until the terminal voltage at A-B is obtained. When performing this calculation, note that if there are resistances in series with the open branch (in this case R_5), they will not affect the evaluation of the potential difference, because no current is flowing through them and they do not drop any voltage.

In order to find the resistance in series with the generator, we must look at the equivalent circuit in *figure 4c*. The individual resistors are combined using the rules discussed earlier for combining resistors in series or parallel. For the calculation of equivalent resistances, the generators are replaced by a short circuit if they are ideal, or by their internal resistance if 'real'.

Having found the two equivalent parameters (voltage and resistance), they can replace the entire network and the effects produced are always equal to those produced by the actual network. In this way, the calculations are simplified considerably.

The superposition theorem

Another principle which can be used in resolving complex electrical circuits is that of **superposition,** which states: the effect of several causes acting together equals the sum of the effects of each cause acting separately. If we consider individual electromotive forces acting on a circuit as a **cause,** and the currents in the individual branches as the **effects,** the individual currents can be obtained as the sum of the partial currents (in each branch) due to the electromotive forces taken separately. Using this method, attention must be paid to the following points:

– when an EMF generator is taken from a network, it is necessary to restore the circuit continuity by replacing the generator with a short-circuit if ideal, or its internal resistance if real;

– the sum of the individual currents is taken as the algebraic sum for which the polarity of the individual magnitudes is taken into account.

Energy conservation in circuit analysis

There are other methods for analysing circuits. One of these is the **principle of the conservation of energy** which is useful, simple and based on a knowledge of electrical power.

This principle states that the total energy of an isolated system remains constant. In other words, the energy of a system may be converted into one form or another (i.e. from potential to kinetic energy), but its total quantity is always the same. This concept applies to electrical power. Firstly, starting with a definition of electrical power, we shall apply it to individual elements of an electrical circuit.

Consider the electrical system illustrated in *figure 5*. The generator in G delivers power at a voltage V and current I over the lines to the 'active user' in U. We have the following quantities of power:

– power from the energy delivered by a generator:

$P_e = E \times I$

where E is the generator EMF and I is the current flowing;

– power corresponding to the energy delivered to the user (P_u). This energy may be provided in another form (e.g. mechanical, thermal etc.). Since it is electrical power it can be represented as a product of voltage and current, so if the active user draws a current I, we have:

$$P_u = V \times I$$

where V is the voltage across the load;

– power corresponding to energy dissipated in the lines:

$$P_p = V_d I = RI^2$$

where V_d is the drop in voltage along the line and R is the resistance of the line.

To balance these quantities of power, it is necessary to define a conventional polarity. We shall consider the power delivered by the generator to be positive, and the power which an electrical system provides externally in any way as negative.

If the principle of the conservation of energy is applied to electrical circuits, we can say that the power generated must equal the power used in any way (including losses on the lines):

$$P_e = P_u + P_p$$

or using the conventional polarities specified above:

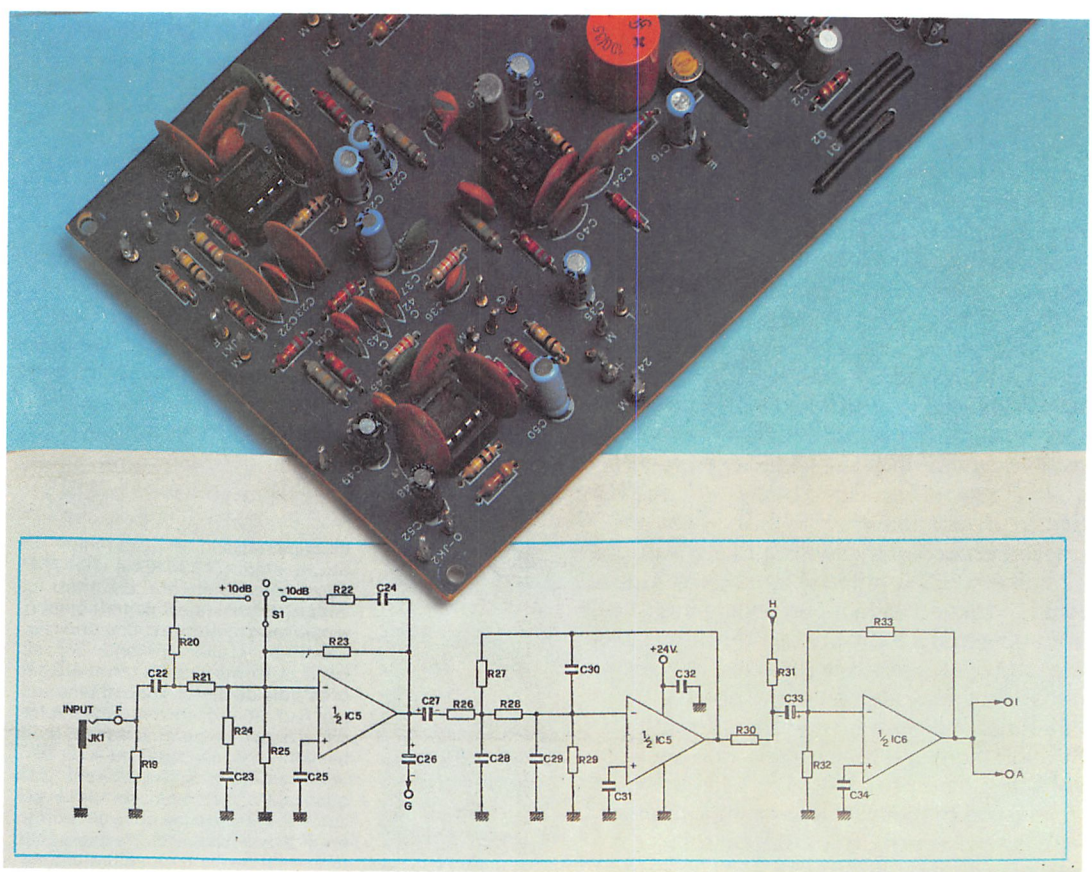
$$P_e - P_u - P_p = 0$$

which, substituting for the individual relationships given, gives:

$$E \times I = (V \times I) + (V_d \times I)$$

In using conventional polarities one can identify the flow of energy by checking whether the current flows from the positive pole (generator or battery) or towards it (from the user). Using the convention means you can evaluate the polarity of the energy handled by a circuit.

The important thing to note is that if there are several quantities of power (delivered or used), they are always totalled. So, if generators are connected in series or in parallel, the corresponding powers are added up to produce the total delivered power. The same applies to lines. The principle of the conservation of energy is therefore always valid. The amount of energy coming in must equal the amount of energy going out, regardless of the connections made in the circuit. □



The diagram and the circuit board of an audio amplifier pre-amp stage. The design and analysis of any electrical circuit relies on the fundamental methods mentioned in this chapter.

ELECTRICAL TECHNOLOGY

Electrical fields

This section explains some of the concepts and definitions needed to understand electrical fields and the forces they generate.

We have seen that insulating materials can hold a charge of static electricity which is capable of attracting small pieces of paper. The phenomenon of electrical force is fundamental to the working of many machines and there is a simple experiment which investigates this further.

An ebonite rod is statically charged by rubbing it with a woollen cloth, it is then hung from a thread so that it remains horizontally balanced and free to rotate. Another similarly charged ebonite rod is moved towards one end of the suspended rod, which will rotate away. This is because both rods are negatively charged, and like charges repel.

However, if we rub a glass rod and move it near to the suspended rod, the ebonite rod will swing towards the glass rod. This is because the glass rod is positively charged and unlike charges attract. The harder the rods are rubbed, the greater the charge will be, thereby increasing the forces of attraction or repulsion. When two opposite charges attract each other, they try to cancel each other out. If bodies with equal numbers of opposite charges come into contact the change on each is neutralized.

These experiments prove the laws of attraction, repulsion and neutralization in charged insulators, but the same laws also apply to charged conductors.

Both conductors and insulators are capable of carrying electrical charges which can vary in size and polarity. The size of the charge is measured in **coulombs** (C). A coulomb is defined as the amount of electrical charge which passes a point in a conductor when a current of 1 A flows for a time of 1 s. This is therefore a measure of the electrons or holes making up the charge.

The force which two charged bodies exert on each other can be calculated by multiplying together the sizes of the charges and dividing the result by the square of the distance between them. The result of this calculation is then multiplied by a factor known as 'permittivity' which accounts for the electrical properties of the material lying between the charges.

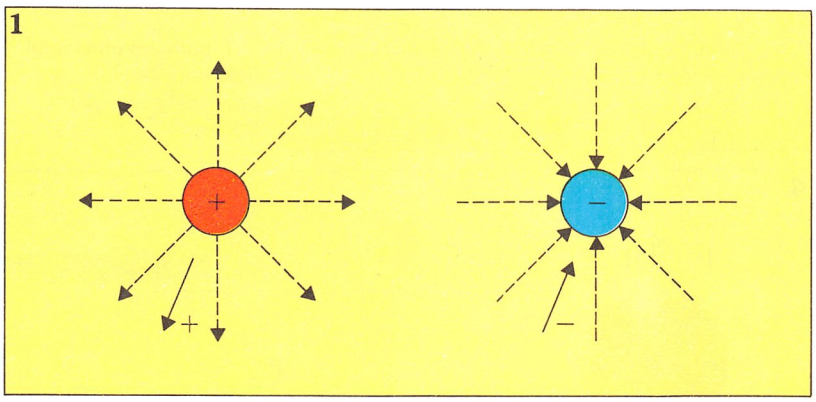
It does not matter how big the conductors or insulators carrying the charge are, or what they are made of – only the size of the charge on them dictates the force they exert on each other.

Imagine a small charged particle of matter floating freely in space. It is surrounded by an invisible region in which any other charged body would feel a force. The closer together these two bodies are then the stronger the force.

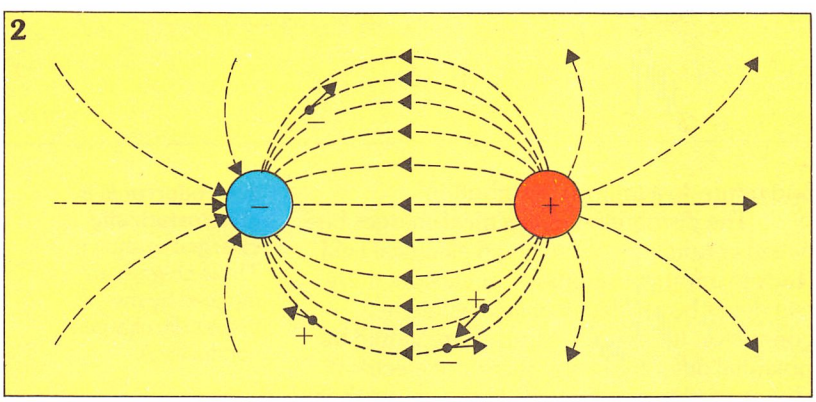
This invisible region is an electrical field which is similar in nature to the space around the earth where a body would feel the effects of gravity. The electrical field is usually illustrated by lines coming out of, or going into the charged body. These are called **lines of force.** Their direction at any point shows the way in which a tiny positively charged body at that point would feel a force, and the proximity of the lines to each other represents the intensity of the field.

Figure 1 shows arrows on the lines of force coming from a positive and negative charge, indicating the direction in which the lines of force travel. You will notice that they always leave from, or arrive at right angles to the surface of a body made of conducting material.

Figure 2 illustrates how lines of force join up when two charged bodies are close to each other. The direction of the force acting on our positively charged particle in four locations is



1. Conventional lines of force for positively and negatively charged particles.



2. Lines of force between two charged bodies of opposing polarities. The lines of force move from the positive to the negative body.

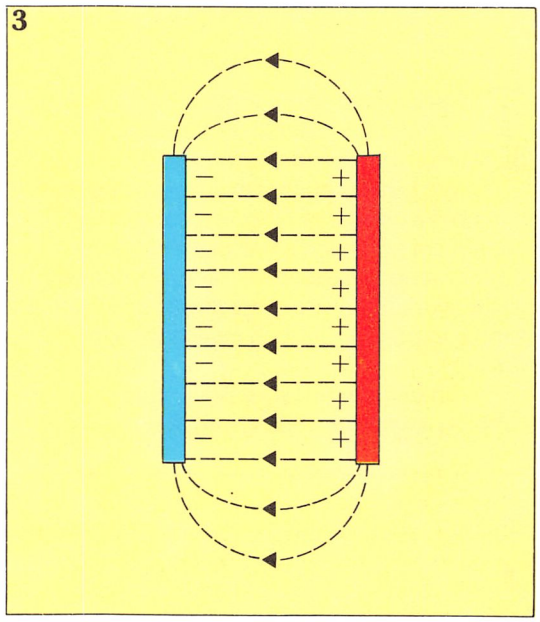
shown. If the charged particle is moved from one point to any other, we have to overcome these forces by doing work. The amount of work done can be calculated by multiplying the size of the charge by the potential difference between the two points. The charge is measured in coulombs, the potential difference in volts and the answer is given in joules, the unit of energy.

Figure 3 shows the effect of the field between two long thin charged plates. The area between the bodies where the lines of force are parallel and equally spaced is called a uniform electrical field. In a uniform electrical field the field intensity is constant.

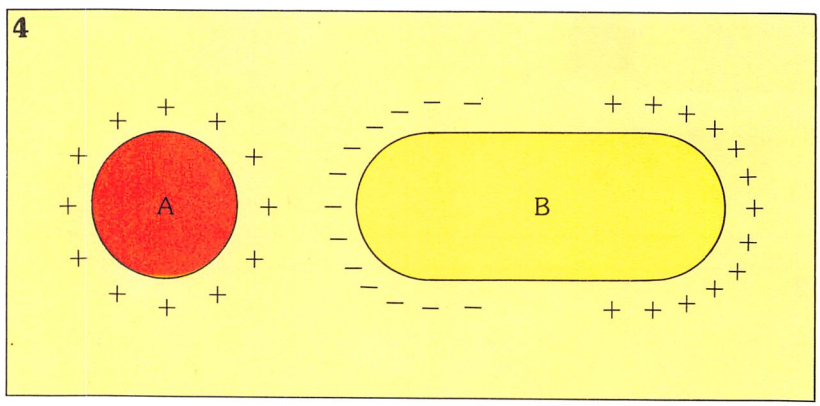
However, the lines of force of an electric field are not always linear, and so potential difference is not always the same between different points. It is therefore necessary to define how the intensity of an electric field varies from one point to another, or how the potential within the field diminishes. Assuming K to be the intensity of the electrical field or electrical force:

$$K = \frac{V}{d}$$

where V is the potential difference between any two points, and d is the distance between them. K is thus measured in volts per metre (Vm^{-1}).



3. A uniform electrical field between two thin, oppositely charged bodies.



4. Conductor B is electrostatically charged as shown. These charges are induced by the previously charged conductor A.

Electrostatic Induction

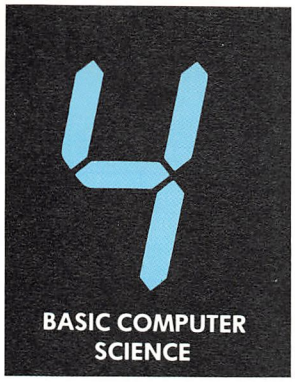
Until now we have assumed that a body can be electrostatically charged by rubbing. A conductor can also be charged by electrostatic induction. *Figure 4* shows an insulated neutral metallic body B, placed in the field of a charged conductor A. The charge on the conductor surface facing A will be of opposite polarity to the charge on A. The charge on the opposite surface of B will be of similar polarity to that on A.

This is because the free electrons of the induced body B are subject to the force of the electrical field and accumulate on the surface at one end, so that the other end becomes positive due to the loss of electrons. If conductor B (charged by induction) is earthed, charges of the same polarity as those of the inducing conductor A will go to earth and B will retain a charge whose polarity is opposite to that of A. If the earth connection is now removed, B will keep this charge even if the conducting body A with its associated field is removed.

Electrostatic induction also occurs on a conductor if it is placed within a uniform electrical field. Within the conductor, the electrons move towards the end which is nearest to the positive pole of the field. Naturally, if the conductor is taken out of the field, the electrons are no longer subject to the force of attraction and return to a uniform distribution.

The distribution of electrical charges between conductors placed in an electrical field leads to another important conclusion: there can never be an electrical field within a conductor. If this wasn't true, there would be potential differences and therefore electrical charges within conductors. These would then be forced to move, thus cancelling the field. So electrical charges always accumulate on the surface of conductors, and occupy a layer which becomes thinner as the potential to which the conductor is subjected is increased.

The concept of zero electrical field within a conductor is used in apparatus like the **Faraday cage.** This is an earthed metallic structure used to protect sensitive equipment from the effects of external electrical fields in a technique called **screening.** □



Memories

Memory characteristics

One of the prime reasons for the success of computers is their ability to memorize instructions and information (data). The devices used for this are known simply as **memories.** This chapter will explain the development of different sorts of memory, how they function, and the reasons why some are more widely used than others.

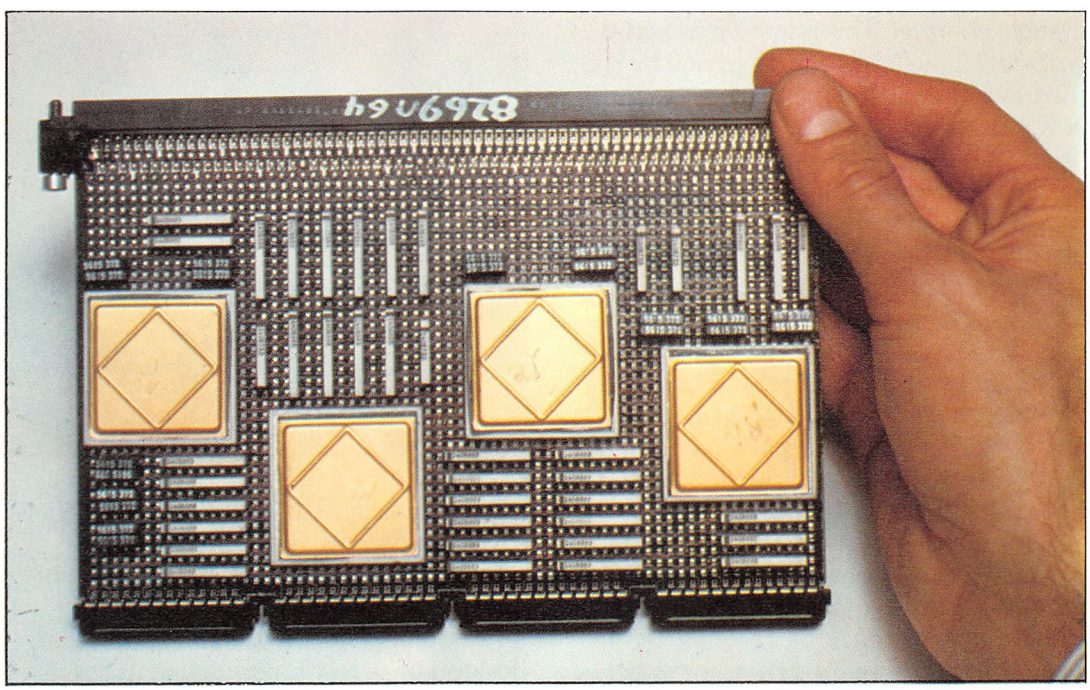
Let's briefly recap on the major concepts of memory which have already been covered.

Memories are information stores which can be divided into individually identifiable elements. In many computers these elements consist of 8 bits (binary digits), or one byte. Other computers divide their memory into words, which may or may not be multiples of bytes. When a word is made up of several bytes, it is usually possible to single out (address) either the individual bytes or the word.

A two-byte word is fairly common. In this case the word can be addressed using even numbered addresses, and the bytes using both odd and even addresses. The address is a numerical identification code which singles out the position of an element within the memory and gives access to it. In some cases, the address is made up of two parts: the first gives access to the memory bank, while the second picks out a single element within that bank.

Memories can be of two basic types: volatile and non-volatile. In a volatile memory, any data held is lost immediately the power supply is cut off. Non-volatile memory, on the other hand, retains all its data even when there is no power supply. The information can only be cancelled on command from the user. In either case, a memory is the device within the computer where data can be deposited, stored, and accessed when required.

Memorization is the process of mod-



A memory board from the IBM 4300. This board is based on microblocks, each of which is able to store 64 K bits on a surface area of 6.35 mm^2 (Photo: courtesy IBM.)

ifying the physical state of the memory in order that it represents the data to be stored. This is achieved by means of a suitable system of coding.

Data transfer

Most computer systems require the transfer of data from one memory to another, so it is worth taking a look at how this is achieved. Should the transfer procedures not be compatible, then a suitable **interface** would be needed to make conversions. Suppose that an item of information is composed of C characters and that n bits are needed for each character. The **total** number of bits (N) to be transferred is $N = C \times n$. The transfer of these bits can take place in any of the following ways:
(1) the bits travel, one after another, along a single channel. This is known as **serial transfer.** If the transfer time of one bit is t, the transfer of all the bits takes place in $N \times t$ units of time;
(2) the n bits making up a character move along n channels simultaneously. This is a **parallel-serial transfer.** It is parallel for the bits of each character, and in series for the sequence of characters. If time t is needed to transfer one character, then the total transfer time is $C \times t$ using n channels;
(3) all the bits of each character of equal weight move along C channels simultaneously. This is **serial-parallel** transfer. Taking t as the transfer time for one bit, the total transfer time is $N \times t$;
(4) all the bits move along n channels in parallel. This is known as **completely parallel transfer** and is the fastest of all, taking place in time t only.

These four methods of transfer are listed in order of increasing speed. The faster the method, the greater the cost of the equipment.

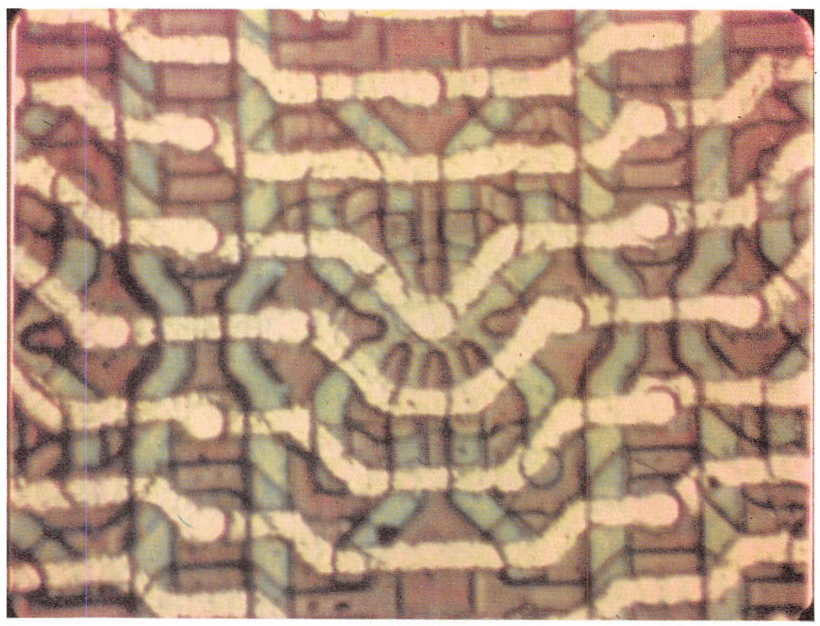
Central memory

The central memory of a computer is continually in action throughout the time the computer is in use, because it carries the instructions which drive the program as well as the data being processed. The partial or final results of processing are also 'written' in the central memory before being sent to the peripheral memories (i.e. disks or tapes). One of the characteristics of central memory is that the access time is fixed for every element (**uniform access**).

We have already mentioned that there are two types of central memory. While a program is running the instructions in use are transferred to the faster part of the memory, known as the **cache memory** or **buffer.** The second, larger part of the memory has a slower access speed.

The addressing, as we have seen, is either by banks or by a single element. In the first case, the single element (byte or word) is located by first addressing the bank, while in the second the element is located directly. The elements which characterise a memory are:
- the overall **capacity,** i.e. the amount of data it can contain. This is expressed in terms of bits, bytes or words. Today's desktop or personal computers usually have memory capacities which are multiples of kilobytes (K), for example 16K, 32K, 48K, 64K
- the form of address
- the capacity of the smallest element which can be addressed individually
- the access time, i.e. the time taken for data requested via the control unit to become available (this is usually expressed in nanoseconds)
- the type of access permitted
- the volatility or otherwise of its data.

We will discuss each type of built-in and peripheral memory in terms of these elements.



Microphotograph of a static RAM cell.

Ferrite core memory

1. The hysteresis loop of a ferrite core.

2. Diagram showing how ferrite cores are wired up to form a memory matrix.

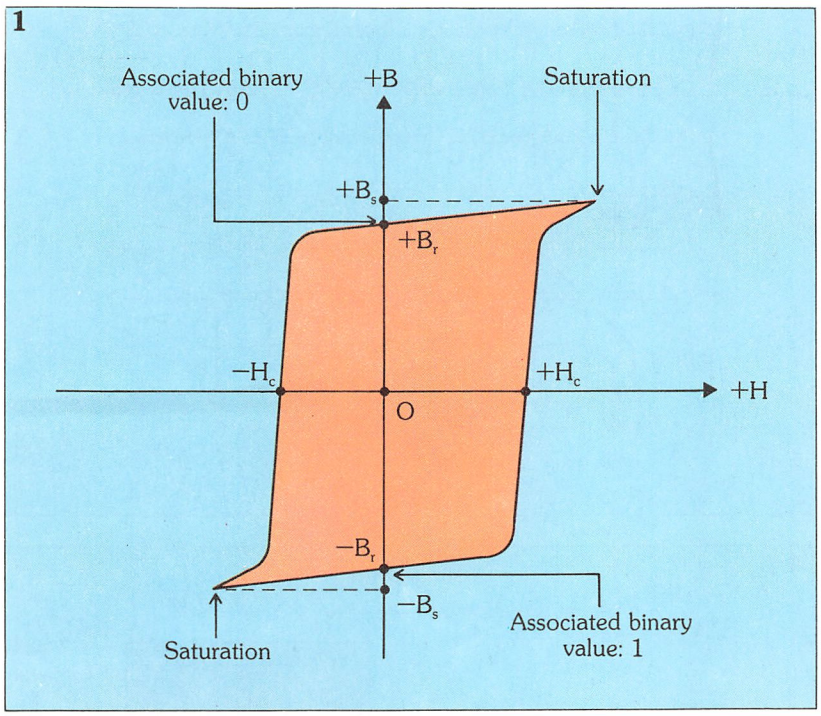
This is the oldest form of magnetic memory. Referring to it as **ferrite core memory** avoids confusion, as these days 'core' is often used to refer to any computer's central memory, whatever type it is. Ferrite core memory is a non-volatile, quick, direct access memory and was used in large scale mainframe computers for a number of years until the mid-1970s. It has now been almost entirely superseded by semiconductor memory.

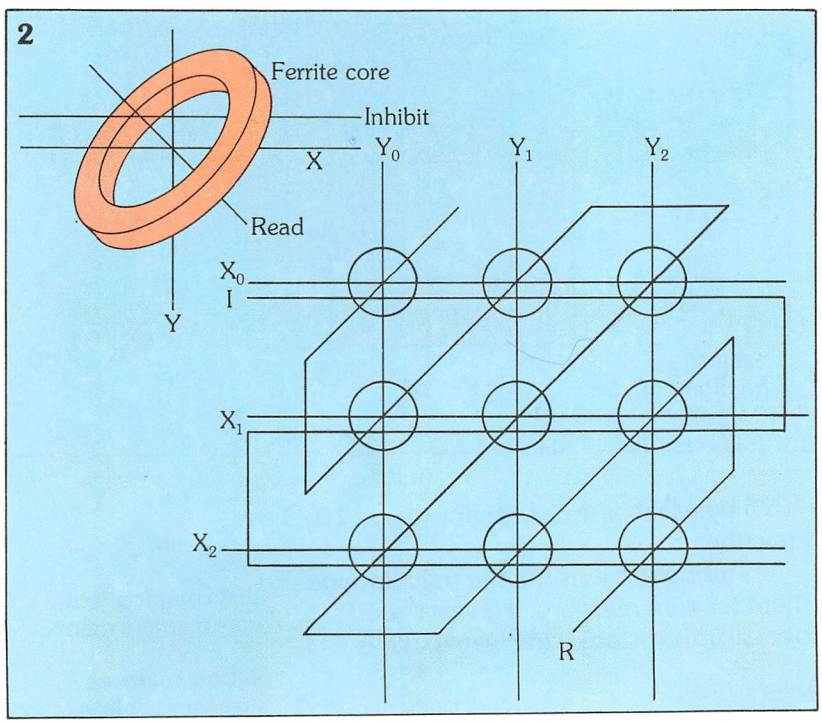
As semiconductor memories cost a fraction of ferrite core memories, take up less space and consume less power they are used extensively in today's computers. However, there are still systems around which use ferrite core, so it is important to get an idea of how it works.

Ferrite, which is composed of iron oxide compressed in a ceramic base so that it can be shaped into any form, is used in core memory because it is easily magnetized. Its hysteresis loop is shown in *figure 1*. The ferrite core memory consists of a matrix of wires with a ferrite ring (core) at each intersection. The direction of magnetism in each core indicates a logic state of 1 or 0 *(see figure 2)*.

Each matrix of cores makes up a plane of memory storage, *figure 3* shows the overall arrangement of such a matrix. *Figure 4* shows how these matrices are arranged in planes. Each group of planes usually corresponds to one byte. To read a byte of data from such an arrangement, one bit would be taken from each matrix in turn.

As we have said, the direction of magnetism in each core is used to represent the binary digits 0 and 1. The direction of magnetism that represents 0 is opposite to that which represents 1.





Storing data in core memory

So, how do we store and read binary digits in ferrite core memory? If you look at *figure 2*, you can see that each individual core has four wires passing through it. They are labelled **X, Y, read** and **inhibit.** These are used to read and write data in the core.

If a matrix of cores is available for data storage, all the ferrite rings are initially magnetised to represent 0. To store the binary digit 1 in a ferrite core, the direction of magnetism present in that core must be changed. This is done by sending an electric current through the core. However if you look at *figure 3* you will see that the wires that pass through the cores are shared. To send a current down one wire would magnetize all the cores that the wire passes through.

So, how do you isolate a particular core? The answer to this lies in the matrix. Each core on the matrix has a particular co-ordinate that is defined by one X wire and one Y wire. If half the current that will magnetize the core is sent down each of the corresponding X and Y wires, then the chosen ferrite ring will receive the full current and will be magnetized.

To record a 1 the current which is sent down each of the two wires is equal to $+\frac{1}{2}$. This changes the direction of magnetism in the core to represent a 1.

To record a 0, the machine still sends two currents of $+\frac{1}{2}$ down the X and Y wires, but counteracts them by sending a current of $-\frac{1}{2}$ down the inhibit wire. This ensures that the core remains in the 0 state, because the total current is only equal to $+\frac{1}{2}$. The current sent down the inhibit wire is not large enough to affect any of the other cores it passes through.

Reading data from core store

To read data from core memory a current of $-\frac{1}{2}$ is sent down the X and Y wires which correspond to the storage position required. This is equal to a total current of $-I$. If the core that is being interrogated is storing a 1, then the current $-I$ will change the direction of magnetism in the core. In changing the direction of the magnetic flux a small current pulse is induced in the read wire which passes through all of the cores. This current pulse is interpreted as a 1 by the computer.

If, on the other hand, a 0 is stored in the core, then the direction of the magnetic flux will **not** change. Consequently there will not be a current pulse in the read wire. The computer reads this as a 0.

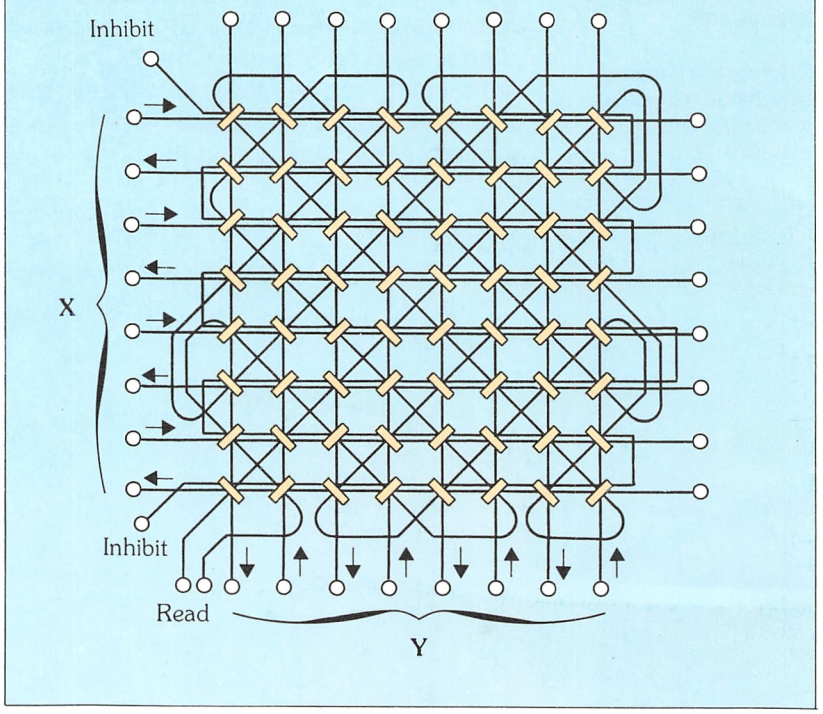
When data is read from core store, the action of reading destroys the contents of the cores. To preserve the contents of the memory each bit must be rewritten after it has been read.

The construction of a core memory is an extremely delicate task because of the tiny dimensions of the cores. They are only millimetres in external diameter and thickness is measured on an even smaller scale. A matrix of 64 × 64 cores, which is capable of registering 4096 bits, is less than 10 cm wide and is about 1 cm thick. The group of 8 matrices necessary to register 4096 bytes takes up less space than a 10 cm cube.

Although this is not the only arrangement for core memory, it gives a good overall impression of the way it works.

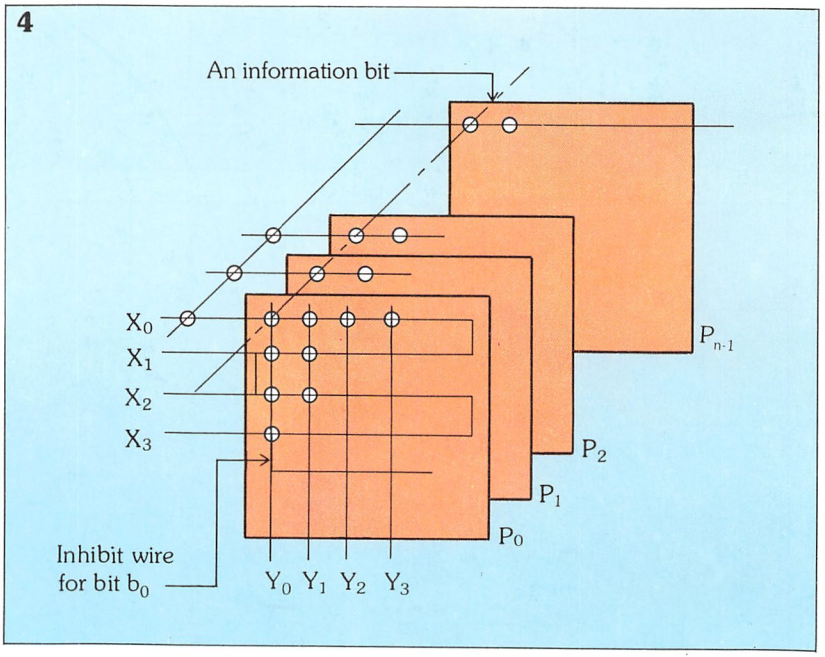
(continued in Part 7)

3



3. A complete ferrite core memory matrix.

4



4. Core matrices arranged in planes.